BEST AVAILABLE COPY

CERTIFICATION OF FACSIMILE TRANSMISSION

I hereby certify that this paper is being facsimile transmitted to The United States Patent and Trademark Office on the date set forth below.
Date Of Signature And Transmission: 6/20/05
Jean C. Baker, Reg. No. 35,433

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Applicants:	Diana Downs, et al.
Serial No.:	09/955,502
Filed:	September 18, 2001
Title:	METHOD FOR PREVENTING SUPEROXIDE DAMAGE TO CELLS AND OXYGEN-LABILE PROTEINS
Group Art Unit:	1645
Examiner:	Patricia Ann Duffy
Docket No.:	960296.97559

COVER PAGE

Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

Dear Sir:

In the matter of the above-identified file, Applicant is faxing the enclosed eight references to Examiner Duffy. These references are Diana M. Downs (Curr. Issues Mol. Biol. 5:17-25, 2003); George M. Weinstock, et al. (Res. Microbiol. 151:151 158, 2000); Joe Handelsman, et al. (Chem. Biol. 5(10):R245-R249, 1998); Michelle R. Rondon, et al. (Proc. Natl. Acad. Sci. USA 96:6451-6455, 1999); Joo-Heon Park, et al. (J. Bacteriol. 186(5):1571-1573, 2004); Jesse D. Woodson and Jorge C. Escalante-Semerena (PNAS 101(10):3591-3596, 2004); Tracey L. Grimek and Jorge C. Escalante-Semerena (J. Bacteriol. 186(2):454-462, 2004); Jesse D. Woodson, et al. (J. Bacteriol. 185(24):7193-7210, 2003).

BEST AVAILABLE COPY

No fee is believed necessary for the Examiner to consider this fax. If a fee is necessary, please charge Deposit Account 17-0055.

Respectfully submitted

Diana M. Downs, et al.

June 21, 2005

By: Jean C. Baker

Jean C. Baker
Attorney for Applicants
Reg. No. 35,433
Quarles & Brady LLP
411 East Wisconsin Avenue
Milwaukee, WI 53202-4497
(414) 277-5709

Curr. Issues Mol. Biol. (2003) 5: 17-25.

Genomics and Bacterial Metabolism

Diana M. Downs*

Department of Bacteriology, University of Wisconsin, Madison, 1550 Linden Dr., Madison, WI 53706 USA

Abstract

The field of bacterial metabolism and physiology is arguably the oldest in microbiology. Much of our understanding of biological processes and molecular paradigms has its roots in early studies of prokaryotic physiology. After a period of declining interest in metabolic studies (prompted by the insurgence of molecular techniques), genomic technologies are revitalizing the study of bacterial metabolism and physiology. These new technologies bring a means to approach metabolic questions with a global perspective. When used in combination with classical and molecular techniques, emerging global technologies will make it feasible to understand the complex integration of metabolic processes that result in an efficient physiology. At the same time, without increased computational capabilities, the massive amounts of data generated by these technologies threaten to overwhelm, rather than facilitate, this work. For genomic technologies to reach their potential for increasing our understanding of bacterial metabolism, microbiologists must become more collaborative and multidisciplinary than at any time in our history.

Introduction

At the core of any living cell is metabolism, the sum total of all the biochemical processes contributing to cell function. Coordination of these processes results in the physiology we associate with each organism, from bacteria to humans. The field of bacterial metabolism and physiology is arguably the oldest in microbiology. Through the years investigators have recognized the need to understand the fundamental principles of life processes and realized the advantages of using prokaryotic cells to achieve this. The field of microbial physiology has changed through the years. With the advent of techniques that allowed researchers to focus more on molecular details of gene structure, organization and expression, research and interest in classical metabolism and physiology declined. Knowledge of complete genome sequences and the ability to visualize the transcription of entire genomes have brought a renewed appreciation of the need to understand the physiology and metabolism of a bacterial cell. Thus a new generation of microbial physiologists is emerging; those with genetic, biochemical, molecular, and genomic tools in their hands. Without a doubt, the application of this powerful combination of approaches to the study of bacterial metabolism and physiology will result in impressive strides in our understanding of basic metabolic processes and their integration. While newly developed technologies will facilitate progress in understanding metabolism, it is critical to keep in mind that the classical approaches, and rigorous definition of biochemical function must not be lost in the mix.

Genomic Techniques Encourage Global Thinking

Since the term was coined by Thomas Roderick in 1987, the field of genomics has undergone extensive growth. Initial skepticism from the scientific community about the validity and reproducibility of global techniques (*e.g.*, expression arrays) has largely been replaced by an appreciation for the value of the data that can be generated. This appreciation remains appropriately tempered by a recognition of the limitations of these techniques. Commentaries and critiques continue to force the field to reassess techniques, approaches, conclusions, and goals (Downs and Escalante-Semerena, 2000; Grunow and Brade, 2001; Perego and Hoch, 2001; Zhou and Miller, 2002). The success of this field can be measured by the impact it has had on investigators in all biological disciplines. The flood of press and data generated by global technologies has facilitated, and in fact encouraged, investigators to realize that each subject they study is also a component of a larger complex system. This realization has brought the field of microbial physiology full cycle. Classical microbial physiologists considered the contributions of a whole system, whether it was a cell or a population of cells, when they measured properties of carbon utilization, nitrogen fixation, etc. The advent of molecular biology brought a new opportunity to understand molecular details of components within the cell. The premise that a detailed knowledge of the components would provide an understanding of the whole, encouraged decades of reductionist studies that have produced a solid understanding of a large number of molecular processes. Now, with the advent/implementation of global technologies, the pendulum is swinging back, and research characterized by a global perspective is increasing. In the same way classical physiology was enhanced by molecular techniques, emerging genomic technologies provide another dimension to the study of bacterial metabolism/physiology. Data from global analyses (expression profiles, protein profiles, etc.) provide a framework to identify correlations and generate hypotheses. The pursuit and rigorous testing of these hypotheses, not the accumulation of data, will characterize the success of metabolic studies in the genomics era.

Technology has advanced to the point that expression profiles, protein profiles, and other global patterns, are routine enough to be used as a "global phenotype". Reports of global analyses (particularly expression arrays) are

---

*For correspondence. Email downs@bact.wisc.edu; Tel. 608-265-4630; Fax. 608-262-9005.

© 2003 Caister Academic Press

increasing exponentially in the literature. From these data new hypotheses about the role of a sequence, the function of a protein or the extent of a regulon, are being generated. These technologies force a broad perspective that previous researchers did not have, and thus provide an additional context in which metabolic dogmas can be tested. When the data analyses from global technologies generate new hypotheses and force old dogmas to be questioned, it provides a healthy infusion into the field of bacterial metabolism and physiology (Gonzalez *et al.*, 2002; Rhodius *et al.*, 2002; Wasserman *et al.*, 2001).

For all the potential of approaches using global techniques, caution must be taken with their use. A tenet long held by biologists is that genes expressed together are utilized together. While a valid hypothesis, this tenet is not a basis to assign function. The number of expression profiles being reported is becoming overwhelming. Among these are global data obtained from numerous bacteria, different growth conditions, and different genetic backgrounds. They represent regulation of gene expression due to nutrient deprivation, antibiotic treatment, and carefully chosen genetic defects (Karlin *et al.*, 2001; Ng *et al.*, 2003, Tao *et al.*, 1999; Weber and Jung, 2002; Zheng *et al.*, 2001). Without question these studies have opened new and exciting directions for research in our efforts to understand gene function and the complexities of bacterial metabolism. The sensitivity of these techniques to small differences in expression levels, and the fact that populations are being analyzed, make it critical that investigators are conscious of subtle changes that might occur in growth conditions or sample preparation. Without careful controls, investigators can misinterpret the cause (and thus the relevance) of the variations in expression detected. In addition, a change in transcript level can reflect either direct or indirect effects of the parameter being addressed. Further, the level of transcription does not always correlate with the level of protein, much less functional protein. For this reason, the global analysis of protein (*i.e.*, proteomics) is gaining popularity (Jungblut *et al.*, 2001).

Global analyses of gene expression are being preformed by investigators from a variety of disciplines with a number of goals. While this use reflects the broad value of these technologies, this diversity and rapid growth make it more difficult to enforce standards in the technology. If genomic technologies are to be used efficiently, efforts must be made to ensure the data from these studies are considered in the context of existing metabolic literature. Relevant research results from the past should be identified and addressed in models that are presented based on genomic data. Maintaining this inclusion becomes difficult as researchers from multiple disciplines, that may not be aware of this literature, employ these technologies. Conclusions about metabolism based on genomic technologies that appear to violate a vast literature, need to be recognized and discussed. When considered in isolation, expression array data is of limited use, but in combination with knowledge from decades of metabolic and regulatory studies, it can generate a deeper understanding of bacterial metabolism and physiology. If the extensive literature from previous metabolic studies is not incorporated into the thinking supported by new technologies, investigators will too often end up "rediscovering the wheel".

The amount of literature one must be aware of to ensure identifying all processes that may impinge on a single pathway in the cell is staggering. A concern with global approaches is that so much data can be generated so quickly, it will be nearly impossible for an investigator to identify and incorporate the literature that he/she should be aware of in interpreting those data. Collaboration and extensive communication between researchers versed in genomic approaches and those with a knowledge of classical physiology and metabolism, in addition to advances in data storage/analysis, will help address this concern.

Model Organisms-Extending the Paradigms

Results from genome sequencing efforts have emphasized that metabolic processes are conserved in diverse living organisms. These results have led to approaches to reconstruct metabolism in diverse organisms *in silico* (Castresana, 2001, Mittenhuber, 2001; Schilling *et al.*, 2002). Perhaps more importantly, this realization has re-invigorated metabolic research in prokaryotes and validated the study of model microorganisms as a means to define metabolic paradigms. In the case of metabolic analyses, one requirement of a model organism is that *in vivo* analyses are possible. Numerous reports appear every month, of work in model organisms (*e.g.*, *E. coli*, *S. enterica*, *B.subtilis*) that has been prompted, facilitated, and/or justified by comparative genomic analyses. Frequently one reads about a locus and/or phenomenon identified in a non-model organism and investigators are quickly drawn to the system in *E. coli* (or other model organism) due to technical ease, and the vast metabolic knowledge base in this bacterium. An example of this scenario is the study of Fe-S cluster biogenesis, a field that has undergone explosive growth in the last ten years. Work by Dean and others on nitrogenase in *A. vinelandii* identified gene products (NifSUA) required for the formation of metal clusters in this complex enzyme system (Jacobson *et al.*, 1989; Zheng *et al.*, 1993). These results were followed by the identification of a similar set of genes (*isc*) elsewhere in the chromosome of *A. vinelandii* (Zheng *et al.*, 1998), and the subsequent demonstration that these genes were conserved from bacteria to humans, and had a critical role in the biogenesis of Fe-S clusters *in vivo* (Campos Garcia and Soberon-Chavez, 2000; Holt *et al.*, 2000; Lill and Kispal, 2000; Mansy *et al.*, 2002; Schwartz *et al.*, 2000; Soldier *et al.*, 2001; Skovran and Downs, 2000; Tachezy *et al.*, 2001; Takahashi and Nakamura, 1999; Voisine *et al.*, 2001). While the biochemistry of the gene products has been pursued in a number of organisms, research on the genetic locus as well as its regulation, and physiological role of the gene products has been most rapid in model organisms (Holt *et al.*, 2000; Kambampati and Lauhon, 2000; Kambampati and Lauhon, 1999; Ollagnier-de-Choudens *et al.*, 2001; Schwartz *et al.*, 2001; Smith *et al.*, 2001; Tokumoto and Takahashi, 2001; Urbina *et al.*, 2001). The emerging characterization of the *suf* genes (recognized now to be at least partially redundant with the isc system) provides more evidence of the productivity that can be obtained by studying the relevant problem in a model organism, and combining these results with the input of data from work in other systems (Patzer and Hantke, 1999; Rangachari *et al.*, 2002). The rapid growth in this, and other, fields of metabolism can be largely attributed the presence of comparative genomic technologies.

A slightly different example involves identification of a locus in the model organism, that has a homolog identified or studied for its role in a metabolic process(es) in a distinct organism. If, as is often the case in genetically tractable organisms, the lesion was identified by a phenotypic screen, the studies of homologs can provide a new perspective for the functional characterization that follows. The recent work on competence gene homologs in *E. coli* by Finkel *et al.* is a good illustration of this scenario. In this case, use of novel phenotypic analyses (*i.e.*, competitiveness in long-term stationary phase), led to the identification of genes required for the ability of *E. coli* to use DNA as a nutrient (Finkel and Kolter, 2001). Although not surprising in retrospect, the genes involved were homologs of genes in *Haemophilus influenzae* and *Neisseria gonorrhoeae* that had been ascribed a role in natural competence (Dougherty and Smith, 1999; Smith *et al.*, 1995). This finding facilitates progress in the understanding of two distinct, but similar metabolic processes. In the absence of comparative genomic technologies, each of the research groups would have a more difficult time identifying a function and physiological role of the gene(s) involved in their respective processes. This example also highlights the benefit that would be derived from knowing the physical identity of genetic loci that have been described in the literature. Such a correlation would allow researchers to take advantage of the physiology and phenotypes that have been described for mutants in multiple systems throughout the years.

While the above examples are directly impacted by the plethora of genome sequences and comparative genomic technologies, arguably the most important work in model organisms is the continuation of the efforts aimed at understanding basic metabolism that have been preformed for decades. This kind of methodical metabolic and molecular work will continue to define metabolic processes and paradigms, thus facilitating work in other systems. One should remember that a major factor in the speed at which genomic analyses of organisms progresses is the base of metabolic knowledge, and molecularly defined paradigms that have arisen as a the result of decades of rigorous biochemistry, genetics, and molecular biology in the model organisms. Model organisms provide the logical forum to continue the molecular characterization of cellular processes, and uncover new metabolic paradigms. Without this work, the risk exists that analysis of sequence data will degenerate into no more than a means to catalog genes and proteins. A mechanism must remain in place to uncover new areas of research and functional paradigms. Genomic and sequencing technologies have not provided a magic bullet for understanding metabolism, they are simply one more tool in the arsenal available to the modern microbial physiologist. The need for solid basic metabolic research in model organisms is unlikely to diminish in the foreseeable future.

Model Organisms Facilitate Genomic Analysis of Diverse Organisms

The use of model organisms goes beyond analysis of their own metabolism. Model organisms are designated as such because they offer technical ease, a property that can be useful in probing the metabolism of other organisms. The annotated genomes of model organisms provide functional information that is more difficult to access in less tractable systems. For instance, complementation of a mutant phenotype in genetically less tractable organisms by an *E. coli* clone of known function can provide insight about the function disrupted in the parent strain. Conversely, genetically defined organisms can be used to identify functional homologs from diverse organisms. Plasmid libraries of DNA from an organism of interest can be generated in a vector that can replicate in *E. coli*. When these plasmids are introduced to the appropriate mutant strain of *E. coli*, the plasmid(s) that complement the defect are candidates for carrying a functional homolog of the protein missing in the *E. coli* mutant (Bull *et al.*, 1994; Pascopella *et al.*, 1994). Use of heterologous systems (often but not always *E. coli*) has become prevalent in identifying genetic or biochemical functions from less tractable systems. The technologies available, in combination with comparative genomic capabilities have almost made *E. coli* a required lab reagent.

Metabolic properties of diverse organisms can be identified and studied by introducing the genetic material required for the relevant metabolism into model host organisms (Handelsman *et al.*, 1998; Rondon *et al.*, 1999; Weinstock *et al.*, 2000). This approach has been championed in the emerging fields of metagenomics (Rondon *et al.*, 2000) and metabolic engineering (Cameron and Chaplen, 1997). In the former, metabolic capabilities (*e.g.*, antibiotic production) can be identified even in the absence of a culturable parent organism. Appropriate host strains are identified based on the metabolic process of interest and other properties. Using this principle, metabolic capabilities can be "mined" from any environment by obtaining, and cloning, heterogenous DNA preparations. Following introduction of the resulting clones into the appropriate organism, a metabolic capability of interest can be identified as a "gained property" of the host organism. In the second, slightly different situation, organisms with properties of interest are used as a source of donor DNA. In this case, the desired metabolic property is often an ability to biosynthesize an antibiotic, or a given compound whose biosynthetic pathway is the target of engineering focus (Trauger and Walsh, 2000). The primary assumptions of these approaches are that, i) the host organism will provide "supporting metabolism" such that what is identified is the genetic information uniquely needed for the respective metabolic property, and ii) the genetic material for this process is located in a single region of the donor chromosome, such that is will be contained on a contiguous piece of DNA. These assumptions are more or less valid depending on the situation, but the ease at which those kinds of approaches can now be applied, have attracted the attention of investigators in academia and industry. After the genetic material is identified from *in vivo* (or *in situ*) sources, it is useful in metabolic engineering efforts, where DNA encoding desired functions can be manipulated and/or mutagenized, to design and modify pathways for commercial gain (Aldor and Keasling, 2001; Aldor *et al.*, 2002; McDaniel *et al.*, 1999; Pfeifer *et al.*, 2001; Pfeifer and Khosla, 2001; Stachelhaus *et al.*, 1995).

In the above cases illustrating the potential for studies with heterologous metabolisms, the limitations of the system should be kept in mind. While these approaches offer a powerful starting point, the integrated metabolism of each organism is the result of years of selective pressures and the subtleties of regulation and efficiency are unlikely to be mimicked in heterologous systems. While manipulating these genomes and organisms to reach a goal or target a component, the benefit of understanding homogeneous systems should be kept in mind and pursued at a later point.

Functional Genomics - Or Classical Genetics?

At some level the goal of all microbial physiologists is to understand the function(s) of each protein and how they work together to result in the efficient physiology we associate with a living cell. The work of most investigators falls somewhere along the continuum of identifying a gene, to understanding its role and interactions *in vivo*. For a number of reasons, few labs strive to span the entire continuum. The more common scenario is when a genomic investigation identifies gene(s) that are connected (*i.e.*, by regulation, location, homology, etc.,) to the area of research focus in a laboratory. In these situations, a putative function for the gene product often exists based on information that led to an interest in this gene. The function of the relevant gene is then pursued in the context of other research in the laboratory. Another powerful means to identify functionally connected components *in vivo* is by using genetic suppressor analysis. This approach will identify genes that are involved in a process of interest, however indirectly (Downs and Escalante-Semerena, 2000; Enos-Berlage and Downs, 1997; Enos-Berlage *et al.*, 1998; Gralnick and Downs, 2001; Petersen *et al.*, 1996). The difficulty in this approach is that genes that are identified by these mutant analyses often have no obvious connection to the process of interest. Thus the investigator may struggle to define the function of the relevant gene product and the explanation for its interaction with the initial process. So, while this approach has significant potential to provide a functional context for unknown genes *in vivo*, few investigators take advantage of it.

Efforts aimed at identifying the function of proteins on a global scale fall under the poorly defined rubric of "functional genomics". Often these approaches depend on a transcriptome, proteome, metabolome or the result of characterization by some other "omic" technique (Eymann *et al.*, 2002; Tao *et al.*, 1999; Tumbula and Whitman; Wen and Burns, 2002). Many times these efforts simply result in a catalog of genes with respect to a given condition or genetic background. While these analyses provide valuable data, they are not a demonstration of function. One must realize that cellular function implies a known role *in vivo*. Understanding the function of a protein in vivo requires that two things be true. First, there must be a biochemical function for the protein that is demonstrable *in vitro*. Second, lack of this biochemical function must explain the phenotype that results in a cell lacking the relevant protein. Hence neither result alone, and certainly not a solely global analysis, will allow definitive annotation of a gene. A common "functional genomic" strategy being pursued in a number of organisms is to knock out all genes in turn, and analyze the resulting phenotype. While this sounds promising at first, the value of this approach is limited. Consider that in some organisms, notably *E. coli/S. enterica*, classical genetic approaches, scoring a variety of phenotypes, have been performed for more that 50 years. Without creatively screening phenotypes, how likely is this strategy to uncover genes that have not been identified? In the best case scenario, this approach may identify a few additional genes whose absence results in a clear growth phenotype. An immediate benefit could be derived from these approaches, if the data were interpreted in the context of past literature to assign a physical location to genetically defined loci in the literature (Dougherty and Downs, 2003; Frodyma and Downs, 1998; Roberts and Reeve, 1970, Sanderson and Roth, 1988). After decades of metabolic genetics and biochemistry in a number of organisms, our understanding of basic metabolism is good. Pushing the understanding of metabolism/physiology to the next level will require that creative approaches be used to address phenotypes caused by lack of additional cellular components. It is reasonable that many of the ORFs that remain completely uncharacterized are not required for a process to function, but rather for its optimization. Other genetic approaches must be considered to uncover the subtle and/or conditional defects associated with loss of these accessory or redundant proteins. Strategies that have been successful in identifying proteins with this type function protein include generating synthetic lethal mutants (or synthetic auxotrophs), scoring reduced not eliminated growth, and multi-copy suppression analyses (Gralnick and Downs, 2001; Petersen and Downs, 1996; Petersen *et al.*, 1998; Petersen and Downs, 1997; Trzebiatowski *et al.*, 1994). Other creative approaches wait to be pioneered.

Even at their best, the scenarios described above will identify a condition where the relevant gene product is needed, a far cry from knowing the *in vivo* biochemical function. The number of proteins whose only homologs are other undefined proteins with a similar motif, suggests we do not have an understanding of all functional classes of proteins. It is possible, and even likely, that these protein families define new functional and biochemical paradigms that have not been characterized. The question becomes how to go past cataloging by sequence similarity and expression studies to identify the biochemical function? This question is at the crux of an ability to extend our understanding of metabolism, and there is no easy answer. Nor is there a computer program with the power to predict function from unique primary sequence. It has been suggested that by determining the crystal structure of these uncharacterized proteins, a putative function will be clear. While this structural genomic approach is being pursued in a number of research groups, it is unclear that it has been successful in identifying a function of a completely novel protein (Eisenstein *et al.*, 2000; Volz, 1999; Xu *et al.*, 1999; Zarembinski *et al.*, 1998). It is more likely that work toward functional elucidation will proceed by meticulous "discovery" science that is facilitated by metabolic genetics and biochemistry. Progress in this work cannot be outlined, nor forced, because the outcome is not predictable. These studies must be pursued with a creative mind and the conviction that new paradigms are yet to be discovered. Investigators successful in this research will use molecular, biochemical, genetic and genomic technologies, and follow their scientific intuition. This strategy has resulted in landmark discoveries in the past, and genomic technology has yet to eliminate the need for this kind of unbiased discovery based research.

Data Management and Dissemination

With genomic technologies have come unprecedented volumes of data that must be cataloged, mined and analyzed. This is a major challenge for computational scientists, and significant efforts are being made to ensure data management keeps pace with progress of genomic technologies (Covert *et al.*, 2001; Edwards *et al.*, 2002; Edwards *et al.*, 2001; Edwards *et al.*, 2002; Krauthammer *et al.*, 2002; Mendes, 2002; Palsson, 2002; Papin *et al.*, 2002; Wagner, 2001). While the challenges of designing software to handle, screen and analyze these data are recognized, there are additional data management and dissemination needs required for research in bacterial metabolism to progress efficiently. These challenges include the need to: i) make data broadly available, ii) rigorously and rapidly update genomic annotations, and iii) define common nomenclature rules. The field of metabolism and physiology is characterized by its integrative nature. Progress in this field, more than most, is dependent on the access to knowledge of numerous pathways/processes. Because information that must be considered in metabolic studies is becoming so diverse and widespread, it is often not identified by current researchers and this slows progress and increases redundancy of effort. Much of the potential for efficient progress promised by global techniques will be lost if the scientific community cannot simply determine what data are there to support one or another conclusion.

Correct and current annotation of sequenced genomes is critical for metabolic studies. As mentioned above, significant work on unknown genes, and new genes in diverse organisms is dictated by sequence similarity to annotated genes in model organisms. If investigators are allowed to describe the function of a gene based solely on similarity to an annotated gene in a different organism, and the respective annotation is incorrect or not definitive, the literature can become compromised in a way that is hard to reverse. It is imperative that biologists be involved in updating the annotations and that the standard for annotation be clearly stated in each case, such that appropriate, defensible, conclusions can be made when sequence similarity is detected.

Post-genomic nomenclature is also a concern for the efficiency of metabolic studies (Fields and Johnston, 2002). Genes have traditionally been named based on phenotypic analysis and this has led to problems when the biochemical function of the gene product is determined (Frodyma and Downs, 1998; Skovran and Downs, 2003; Trzebiatowski *et al.*, 1994). A better solution is the current trend to name genes based on chromosomal location (*i.e.*, YXX, or ETM) until a biochemical function can be attributed to the gene product (K. Sanderson and M. Berlyn, personal communication). This is not yet a perfect solution, since, in *E. coli*, both a "b" designation, and the "yxx" designation exist and are used. Unfortunately, the literature has been permeated by cases where multiple names for the same gene are used. Not only is this is a problem that is tedious and time consuming to fix, it is unclear whose domain such an effort should fall in. An extension of this issue is the realization that a large amount of useful phenotypic information is present in the literature from the time when loci were genetically, but not physically defined. If a genetic/physical correlation of these genes were generated, it would allow data from the past to be better incorporated in the context of current work, which is preformed with the mindset of physical location. Unfortunately, in both of the above cases, work to clarify the literature, and eliminate redundancies is unlikely to be preformed on more than a case by case basis unless a major emphasis is put on solving these data management challenges.

In addition to the rather mundane cases described above, there is a growing need for new computational technology to integrate information present in the literature. For metabolic research to move forward productively researchers must be able to consider their results in the context of what is known. We are now in an exciting time where data integration is possible and global thinking is encouraged. Some researchers can be well versed in the connections that exist in a small area of metabolism, but as the understanding of metabolic connections and integration expands, it becomes less feasible for this information to be stored by single investigators.

From the perspective of bacterial metabolism, the challenge facing the computational scientists is to design software that can, i) integrate the information in the literature on individual pathways and processes to present an accessible picture of the metabolic connections that exist *in vivo*, and ii) scan literature for words or phrases that may be buried within a manuscript of a different focus. In the former case, storage of known connections in a form that can be accessed, modeled, and expanded is needed. Some of this work is beginning, and several databases built with this goal in mind exist (Goesmann *et al.*, 2002; Karp, 2001; Karp *et al.*, 2002; Karp *et al.*, 2002). As our progress in understanding metabolic integration continues, there should be some mechanism to incorporate these basic research results into the metabolic databases. Display of this information in a useful way will require some creativity. Both direct biochemical connections, and more indirect connections, which can also affect function of a pathway/process, need to be presented. Importantly, these databases need a clear description of which connections/

activities are documented by experimental data and which are simply hypothesized.

The second case reflects a need for investigators to find references to a particular metabolic process from the literature when the relevant statement may be a small piece of a manuscript. Classical approaches to metabolic characterization often result in valuable observations, which appear peripheral to the original focus of the project. The significance of those peripheral observations may not be apparent to the authors of the original study. Yet, to investigators focusing on a different area of metabolism, these observations may provide critical insight. In such a scenario, the "side" observation might never be accessed by the relevant researchers, since as a minor statement in the manuscript, it would be invisible to standard literature searches. An ability to search the literature for words and phrases that indicate a metabolic connection would facilitate the integrative thinking that investigators in metabolic studies need. Progress in metabolic studies depends on astute researchers recognizing phenotypes and observations that suggest a key connection or explain a metabolic behavior. The difficulty in the field has been communicating these observations in a way that the relevant person(s) has access to it in their thinking. This area cries out for collaboration and communication between experts in a number of disciplines, and identifies an area that could have an enormous impact on the progress of our understanding of bacterial metabolism. When this concept becomes a reality, perhaps the scientific community as a whole will be more receptive to the value of describing phenomena for which a biochemical explanation is not yet readily available.

Conclusions

These are exciting times for the field of bacterial metabolism and physiology. It has long been recognized that the cell was comprised of many integrated metabolic pathways and processes. As our knowledge of individual components has increased, we have reached the point where it is no longer enough to consider a single pathway (or process) without regard its integration with the other cellular components. As global technologies and data management strategies continue to be developed, the capacity to present and catalog the integration of multiple pathways and processes will reach a level not possible for the human brain to retain. Those capabilities will push our understanding of bacterial metabolism to a higher level of complexity.

Ironically, what seems to draw researchers to genomic analyses, *i.e.*, the volume of data reflecting activity in the whole cell, is similar to the characteristics that often drive investigators from the more classical approaches to physiology and metabolism. The methodical "discovery" science of classical metabolic genetics and biochemistry is often characterized by slow paced, weaving progress with numerous wrong turns, required to generate and discard models that can explain complex phenotypes in the context of the cellular biochemistry. The field of bacterial metabolism is characterized by a self-imposed mandate to consider metabolic processes in the context of cellular physiology, rather than in isolation. The difficulty these researchers face is the volume of data, and knowledge of diverse metabolic pathways they must be able to draw from to generate an inclusive model. Computational science has the potential to ease this difficulty in the future and facilitate the integration of data generated from multidisciplinary approaches to metabolism.

Genomic technologies and data analysis techniques are powerful tools to probe bacterial metabolism but they have yet to replace the curiosity and creativity of the human mind in making progress. These technologies have allowed the rapid cataloging of sequence similarities and differences, common regulatory themes and protein stability. The challenge in the field of physiology and metabolism in the future will be to take advantage of these technologies without losing an open mind and the drive to pursue the exceptions, or the places that do not fit existing paradigms. Thus, in addition to genomic technologies, creativity and persistence will be required if we are to take our knowledge of cellular processes and understanding of metabolic integration to the next level by continuing to define biological paradigms.

Acknowledgements

D.M.D. is supported by National Institutes of Health grant GM47296, National Sciences Foundation grant MCB9723830, and a 21st Century Scientist Award from the J.S. McDonnell Foundation.

References

Aldor, I., and Keasling, J.D. 2001. Metabolic engineering of poly(3-hydroxybutyrate-co-3-hydroxyvalerate) composition in recombinant *Salmonella enterica* serovar typhimurium. Biotechnol Bioeng 76: 108-114.

Aldor, I.S., Kim, S.W., Prather, K.L., and Keasling, J.D. 2002. Metabolic engineering of a novel propionate-independent pathway for the production of poly(3-Hydroxybutyrate-co-3-hydroxyvalerate) in recombinant *Salmonella enterica* serovar Typhimurium. Appl. Environ. Microbiol. 68: 3848-3854.

Bull, C.T., Ishimaru, C.A., and Loper, J.E. 1994. Two Genomic Regions Involved in Catechol Siderophore Production by *Erwinia carotovora*. Appl. Environ. Microbiol. 60: 662-669.

Cameron, D.C., and Chaplen, F.W. 1997. Developments in metabolic engineering. Curr. Opin. Biotechnol. 8: 175-180.

Campos-Garcia, J., Ordóñez, L.G., and Soberon-Chavez, G. 2000. The *Pseudomonas aeruginosa hscA* gene encodes Hsc66, a DnaK homologue. Microbiol. 146: 1429-1435.

Castresana, J. 2001. Comparative genomics and bioenergetics. Biochim. Biophys. Acta. 1506: 147-162.

Covert, M.W., Schilling, C.H., Famili, I., Edwards, J.S., Goryanin, I.I., Selkov, E., and Palsson, B.O. 2001. Metabolic modeling of microbial strains *in silico*. Trends Biochem. Sci. 26: 179-186.

Dougherty, B.A., and Smith, H.O. 1999. Identification of *Haemophilus influenzae* Rd transformation genes using cassette mutagenesis. Microbiol. 145: 401-409.

Dougherty, M., and Downs, D.M. 2003. The *stm4066* gene product of *Salmonella enterica* serovar Typhimurium has aminoimidazole riboside (AIRs) kinase activity, and allows AIRs to satisfy the thiamin requirement of *pur* mutant strains. J. Bacteriol. 185: 332-339.

Downs, D.M., and Escalante-Semerena, J.C. 2000. Impact of genomics and genetics on the elucidation of bacterial metabolism. Methods. 20: 47-54.

Edwards, J.S., Covert, M., and Palsson, B. 2002. Metabolic modelling of microbes: the flux-balance approach. Environ. Microbiol. 4: 133-140.

Edwards, J.S., Ibarra, R. U., and Palsson, B.O. 2001. *In silico* predictions of *Escherichia coli* metabolic capabilities are consistent with experimental data. Nat. Biotechnol. 19: 125-130.

Edwards, J.S., Ramakrishna, R., and Palsson, B.O. 2002. Characterizing the metabolic phenotype: a phenotype phase plane analysis. Biotechnol. Bioeng. 77: 27-36.

Eisenstein, E., Gilliland, G.L., Herzberg, O., Moult, J., Orban, J., Poljak, R.J., Banerjei, L., Richardson, D., and Howard, A.J. 2000. Biological function made crystal clear-annotation of hypothetical proteins via structural genomics. Curr. Opin. Biotechnol. 11: 25-30.

Enos-Berlage, J.L., and Downs, D.M. 1997. Mutations in *sdh* (succinate dehydrogenase genes) alter the thiamine requirement of *Salmonella typhimurium*. J. Bacteriol. 179: 3989-3996.

Enos-Berlage, J.L., Langendorf, M.J., and Downs, D.M. 1998. Complex metabolic phenotypes caused by a mutation in *yjgF*, encoding a member of the highly conserved YER057c/YjgF family of proteins. J. Bacteriol. 180: 6519-6528.

Eymann, C., Homuth, G., Scharf, C., and Hecker, M. 2002. *Bacillus subtilis* functional genomics: global characterization of the stringent response by proteome and transcriptome analysis. J. Bacteriol. 184: 2500-2520.

Fields, S., and Johnston, M. 2002. Genomics. A crisis in postgenomic nomenclature. Science. 296: 671-672.

Finkel, S.E., and Kolter, R. 2001. DNA as a nutrient: novel role for bacterial competence gene homologs. J. Bacteriol. 183: 6288-6293.

Frodyma, M., and Downs, D.M. 1998. The *panE* gene, encoding ketopantoate reductase, maps at 10 minutes and is allelic to *apbA* in *Salmonella typhimurium*. J. Bacteriol. 180. 4757-4759.

Goesmann, A., Haubrock, M., Meyer, F., Kalinowski, J., and Giegerich, H. 2002. PathFinder: reconstruction and dynamic visualization of metabolic pathways. Bioinformatics. 18: 124-129.

Gonzalez, R., Tao, H., Shanmugam, K.T., York, S.W., and Ingram, L.O. 2002. Global gene expression differences associated with changes in glycolytic flux and growth rate in *Escherichia coli* during the fermentation of glucose and xylose. Biotechnol. Prog. 18: 6-20.

Gralnick, J., and Downs, D. 2001. Protection from superoxide damage associated with an increased level of the YggX protein in *Salmonella enterica*. Proc. Natl. Acad. Sci. USA. 98: 8030-5.

Gronow, S., and Brade, H. 2001. Lipopolysaccharide biosynthesis: which steps do bacteria need to survive? J. Endotoxin. Res. 7: 3-23.

Handelsman, J., Rondon, M.R., Brady, S.F., Clardy, J., and Goodman, R.M. 1998. Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products. Chem. Biol. 5: R245-249.

Hoff, K.G., Silberg, J.J., and Vickery, L.E. 2000. Interaction of the iron-sulfur cluster assembly protein IscU with the Hsc66/Hsc20 molecular chaperone system of *Escherichia coli*. Proc. Natl. Acad. Sci. USA. 97: 7790-7795.

Jacobson, M.R., Cash, V.L., Weiss, M.C., Laird, N.F., Newton, W.E., and Dean, D.R. 1989. Biochemical and genetic analysis of the *nifUSVWZM* cluster from *Azotobacter vinelandii*. Mol. Gen. Genet. 219: 49-57.

Jungblut, P.R., Muller, E.C., Mattow, J., and Kaufmann, S.H. 2001. Proteomics reveals open reading frames in *Mycobacterium tuberculosis* H37Rv not predicted by genomics. Infect. Immun. 69: 5905-5907.

Kambampati, R., and Lauhon, C.T. 2000. Evidence for the transfer of sulfane sulfur from IscS to ThiI during the in vitro biosynthesis of 4-thiouridine in *Escherichia coli* tRNA. J. Biol. Chem. 275: 10727-10730.

Kambampati, R., and Lauhon, C.T. 1999. IscS is a sulfurtransferase for the in vitro biosynthesis of 4-thiouridine in *Escherichia coli* tRNA. Biochem. 38: 16561-16568.

Karlin, S., Mrazek, J., Campbell, A., and Kaiser, D. 2001. Characterizations of highly expressed genes of four fast-growing bacteria. J. Bacteriol. 183: 5025-5040.

Karp, P.D. 2001. Pathway databases: a case study in computational symbolic theories. Science. 293: 2040-2044.

Karp, P.D., Riley, M., Paley, S.M., and Pellegrini-Toole, A. 2002. The MetaCyc Database. Nucleic Acids Res. 30. 59-61.

Karp, P.D., Riley, M., Saier, M., Paulsen, I.T., Collado-Vides, J., Paley, S.M., Pellegrini-Toole, A., Bonavides, C., and Gama-Castro, S. 2002. The EcoCyc database. Nucleic Acids Res. 30: 56-58.

Krauthammer, M., Kra, P., Iossifov, I., Gomez, S.M., Hripcsak, G., Hatzivassiloglou, V., Friedman, C., and Rzhetsky, A. 2002. Of truth and pathways: chasing bits of information through myriads of articles. Bioinformatics. 18: S249-S257.

Lill, R., and Kispal, G. 2000. Maturation of cellular Fe-S proteins: an essential function of mitochondria. Trends Biochem. Sci. 25: 352-356.

Mansy, S.S., Wu, G., Surerus, K.K., and Cowan, J.A. 2002. Iron-sulfur cluster biosynthesis: *Thermatoga maritima* IscU is a structured iron-sulfur cluster assembly protein. J. Biol. Chem. 4: 4.

McDaniel, R., Thamchaipenet, A., Gustafsson, C., Fu, H., Betlach, M., and Ashley, G. 1999. Multiple genetic modifications of the erythromycin polyketide synthase to produce a library of novel "unnatural" natural products. Proc. Natl. Acad. Sci. USA. 96: 1846-1851.

Mendes, P. 2002. Emerging bioinformatics for the metabolome. Brief Bioinformatics. 3: 134-145.

Mittenhuber, G. 2001. Phylogenetic analyses and comparative genomics of vitamin B6 (pyridoxine) and pyridoxal phosphate biosynthetic pathways. J. Mol. Microbiol. Biotechnol. 3: 1-20.

Ng, W.L., Kazmierczak, K.M., Robertson, G.T., Gilmour, R., and Winkler, M.E. 2003. Transcription Regulation and Signature Patterns Revealed by Microarray Analyses of *Streptococcus pneumoniae* R6 challenged with Sub-Lethal Concentrations of Translation Inhibitors. J. Bacteriol. In press.

Ollagnier-de-Choudens, S., Mattioli, T., Takahashi, Y., and Fontecave, M. 2001. Iron-sulfur cluster assembly: characterization of IscA and evidence for a specific and functional complex with ferredoxin. J. Biol. Chem. 276: 22604-22607.

Palsson, B. 2002. In silico biology through "omics". Nat. Biotechnol. 20: 649-650.

Papin, J.A., Price, N.D., Edwards, J.S., and Palsson, B.B. 2002. The genome-scale metabolic extreme pathway structure in *Haemophilus influenzae* shows significant network redundancy. J. Theor. Biol. 215: 67-82.

Pascopella, L., Collins, F.M., Martin, J.M., Lee, M.H., Hatfull, G.F., Stover, C.K., Bloom, B.R., and Jacobs, W.R. Jr. 1994. Use of in vivo complementation in *Mycobacterium tuberculosis* to identify a genomic fragment associated with virulence. Infect. Immun. 62: 1313-1319.

Patzer, S.I., and Hantke, K. 1999. SufS is a NifS-like protein, and SufD is necessary for stability of the [2Fe-2S] FhuF protein in *Escherichia coli*. J. Bacteriol. 181: 3307-3309.

Perego, M., and Hoch, J.A. 2001. Functional Genomics of Gram-Positive Microorganisms: review of the meeting, San Diego, California, 24 to 28 June 2001. J. Bacteriol. 183: 6973-6978.

Peterson, L., and Downs, D.M. 1996. Mutations in *apbC* (*mrp*) prevent function of the alternative pyrimidine biosynthetic pathway in *Salmonella typhimurium*. J. Bacteriol. 178: 5676-5682.

Petersen, L., Enos-Berlage, J., and Downs, D.M. 1996. Genetic analysis of metabolic crosstalk and its impact on thiamine synthesis in *Salmonella typhimurium*. Genetics. 143: 37-44.

Petersen, L.A., and Downs, D.M. 1997. Identification and characterization of an operon in *Salmonella typhimurium* D.M. involved in thiamine biosynthesis. J. Bacteriol. 179: 4894-4900.

Pfeifer, B.A., Admiraal, S.J., Gramajo, H., Cane, D.E., and Khosla, C. 2001. Biosynthesis of complex polyketides in a metabolically engineered strain of *E. coli*. Science. 291: 1790-1792.

Pfeifer, B.A., and Khosla, C. 2001. Biosynthesis of Polyketides in Heterologous Hosts. Microbiol. Mol. Biol. Rev. 65: 106-118.

Rangachari, K., Davis, C.T., Eccleston, J.F., Hirst, E.M., Saldanha, J.W., Strath, M., and Wilson, R.J. 2002. SufC hydrolyzes ATP and interacts with SufB from *Thermotoga maritima*. FEBS Lett. 514: 225-228.

Rhodius, V., Van Dyk, T.K., Gross, C., and LaRossa, R.A. 2002. Impact of genomic technologies on studies of bacterial gene expression. Ann. Rev. Microbiol. 56: 599-624.

Roberts, L.M., and Reeve, E.C.R. 1970. Two mutations giving low-level streptomycin resistance in *Escherichia coli* K12. Genet. Res. 16: 359-365.

Rondon, M.R., August, P.R., Bettermann, A.D., Brady, S.F., Grossman, T.H., Liles, M.R., Loiacono, K.A., Lynch, B.A., MacNeil, I.A., Minor, C., Tiong, C.L., Gilman, M., Osburne, M.S., Clardy, J., Handelsman, J., and Goodman, R.M. 2000. Cloning the soil metagenome: a strategy for accessing the genetic and functional diversity of uncultured microorganisms. Appl. Environ. Microbiol. 66: 2541-2547.

Rondon, M.R., Raffel, S.J., Goodman, R.M., and Handelsman, J. 1999. Toward functional genomics in bacteria: analysis of gene expression in Escherichia coli from a bacterial artificial chromosome library of Bacillus cereus. Proc. Natl. Acad. Sci. USA. 96: 6451-6455.

Sanderson, K.E., and Roth, J.R. 1988. Linkage map of *Salmonella typhimurium*, Edition VII. Microbiol. Rev. 52: 485-532.

Schilling, C.H., Covert, M.W., Famili, I., Church, G.M., Edwards, J.S., and Palsson, B.O. 2002. Genome-scale metabolic model of *Helicobacter pylori* 26695. J. Bacteriol. 184: 4582-4593.

Schwartz, C.J., Djaman, O., Imlay, J.A., and Kiley, P.J. 2000. The cysteine desulfurase, IscS, has a major role in in vivo Fe-S cluster formation in *Escherichia coli*. Proc. Natl. Acad. Sci. USA. 97: 9009-9014.

Schwartz, C.J., Giel, J.L., Patschkowski, T., Luther, C., Ruzicka, F.J., Beinert, H., and Kiley, P.J. 2001. IscR, an Fe-S cluster-containing transcription factor, represses expression of *Escherichia coli* genes encoding Fe-S cluster assembly proteins. Proc. Natl. Acad. Sci. USA. 98: 14895-14900.

Seidler, A., Jaschkowitz, K., and Wollenberg, M. 2001. Incorporation of iron-sulphur clusters in membrane-bound proteins. Biochem. Soc. Trans. 29: 418-421.

Skovran, E., and Downs, D.M. 2003. Lack of ApbC or ApbE proteins results in a defect in Fe-S cluster Metabolism in *Salmonella enterica* serovar Typhimurium. J. Bacteriol. 185: 98-106.

Skovran, E., and Downs, D.M. 2000. Metabolic defects caused by mutations in the *isc* gene cluster in *Salmonella enterica* serovar Typhimurium: implications for thiamine synthesis. J. Bacteriol. 182: 3896-3903.

Smith, A.D., Agar, J.N., Johnson, K.A., Frazzon, J., Amster, I.J., Dean, D.R., and Johnson, M.K. 2001. Sulfur transfer from IscS to IscU: the first step in iron-sulfur cluster biosynthesis. J. Am. Chem. Soc. 123: 11103-11104.

Smith, H.O., Tomb, J.F., Dougherty, B.A., Fleischmann, R.D., and Venter, J.C. 1995. Frequency and distribution of DNA uptake signal sequences in the *Haemophilus influenzae* Rd genome. Science. 269: 538-540.

Stachelhaus, T., Schneider, A., and Marahiel, M.A. 1995. Rational design of peptide antibiotics by targeted replacement of bacterial and fungal domains. Science. 269: 69-72.

Tachezy, J., Sanchez, L.B., and Muller, M. 2001. Mitochondrial type iron-sulfur cluster assembly in the amitochondriate eukaryotes Trichomonas vaginalis and Giardia intestinalis, as indicated by the phylogeny of IscS. Mol. Biol. Evol. 18: 1919-28.

Takahashi, Y., and Nakamura, M. 1999. Functional assignment of the *ORF2-iscS iscU-iscA-hscB-hscA-fdx-ORF3* gene cluster involved in the assembly of Fe-S clusters in *Escherichia coli*. J. Biochem. (Tokyo) 126: 917-20.

Tao, H., Bausch, C., Richmond, C., Blattner, F.R., and

Res. Microbiol. 151 (2000) 151−158
© 2000 Éditions scientifiques et médicales Elsevier SAS. All rights reserved
S092325080000115-7/REV

From microbial genome sequence to applications

George M. Weinstock[a,b*], David Smajs[a,b], John Hardham[a,c], Steven J. Norris

[a] *Department of Microbiology and Molecular Genetics, University of Texas–Houston Medical School, 6431 Fannin Street, Houston, TX 77030, USA*
[b] *Center for the Study of Emerging and Re-emerging Pathogens, University of Texas–Houston Medical School, 6431 Fannin Street, Houston, TX 77030, USA*
[c] *Department of Pathology and Laboratory Medicine, University of Texas–Houston Medical School, 6431 Fannin Street, Houston, TX 77030, USA*

Abstract — Whole genome sequences of microbial pathogens present new opportunities for clinical applications. Chief among these are development of antimicrobials, diagnostics, and vaccines. While antimicrobial development is a more difficult, long-term prospect, new diagnostics and vaccines are likely to be the first products of microbial genomics. To take advantage of whole genome sequences, methods for production of gene products in surrogate hosts (heterologous expression) are required that will work for large-scale, high-throughput gene expression. This will allow genomic information from even the most experimentally difficult pathogens to be mined for applications. In addition, screening methods to test gene products for their potential as vaccine candidates are needed for large scale screening. These areas for technological development should be stimulated by the potential for converting genomic sequence information into applications. © 2000 Éditions scientifiques et médicales Elsevier SAS

syphilis / vaccine / diagnostics

1. From whole genome sequence to applications

A major goal of microbial genome projects is the development of diagnostics, vaccines, and therapeutics. With the completion of a number of whole genome sequences (WGSs) from pathogenic microorganisms, there are numerous efforts underway to mine these genomes for such applications. In this brief review, we discuss the issues involved in this process, with specific reference to the genome of *Treponema pallidum* [4], causative agent of syphilis, for example.

Perhaps the most difficult of the three applications is the development of new antimicrobial therapeutics. This requires as a first step the identification of targets that are essential to survival of the microbe during infection. WGSs impact this in several ways. First, genes that are essential and single copy in some organisms may be duplicated in other genomes, making them less attractive as targets. This can be readily observed from inspection of a WGS and these genes can be avoided in antimicrobial development. Second, homologs to genes known to be essential in other microbes can be identified in a WGS and, while this does not prove that the gene is essential, it allows it to be cloned and studied or mutated. This latter test is the gold standard for proving a gene to be essential but is complicated since the phenotype of a null mutation in an essential gene is lethality. Since such mutants cannot be recovered directly, they are not isolated in standard mutant hunts and require sophisticated directed genetic approaches that are made much easier by knowing the gene's sequence. This, however, is the easiest part of developing antimicrobials. Once the target is identified the search for

---

* Correspondence and reprints
Tel.: +1 713 500 6083; fax: +1 713 500 5499;
georgew@utmmg.med.uth.tmc.edu

chemicals that inhibit the target's function begins, requiring large chemical libraries, chemical modification of promising molecules to improve efficacy and reduce toxicity to the host, and further studies in animals.

Because of this long road to a new antimicrobial, these applications are not likely to be realized soon. Rather, the other two goals, diagnostics and vaccines, are much more likely to be the first fruits from genomics. These products are relatively less demanding than developing antimicrobials. Furthermore, converting genomic information into these products has the potential for high-throughput approaches, allowing rapid, exhaustive screening of all gene products predicted from a genome. High-throughput approaches are also possible in identifying antimicrobial targets, namely genes that are essential as defined by knock-out mutations [1, 5]. But the subsequent development of chemical inhibitors is more complex than the steps needed in diagnostic or vaccine development.

The steps of a high-throughput pipeline to identifying diagnostic or vaccine candidates are shown in *figure 1*. Two key elements are heterologous expression and primary assays. Since many pathogens are difficult to grow or lack systems for genetic manipulation by the experimentalist, it is often essential to express each gene product in *Escherichia coli* or another suitable surrogate host. A classic example of this is *Treponema pallidum*. Although *T. pallidum* was identified as the cause of syphilis in 1905 [8, 9], it is still not possible to continuously culture this bacterium in the laboratory. Rather it must be propagated in animal hosts, usually rabbit testes. Thus, studies of the gene products of *T. pallidum* require that they be produced in a surrogate host.

Once each gene in an organism can be produced and studied separately, it is then necessary to have some screen for utility. For diagnostics, the antigenicity of each gene product is readily determined using patient serum

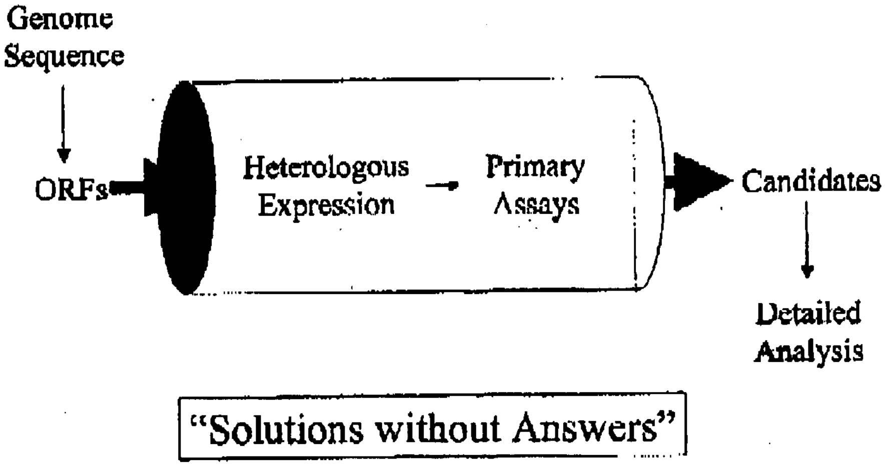


Figure 1. From genome to applications: high-throughput pipeline. Genome sequences are analyzed to predict coding sequences ('ORFs') which can then be cloned into expression vectors to produce each protein encoded by the genome in a surrogate host, such as E. coli, to make subsequent purification more feasible. Each of these gene products can then be tested, in a simple primary assay, for antigenicity in the case of a diagnostic, or for some aspect of protection against challenge in the case of a vaccine. This narrows the list of candidates from 1 000 or more in the initial genome screen to a more manageable number that can then be studied in detail.

samples. Testing for immunogenicity is more difficult since one cannot necessarily set up an elaborate, rigorous vaccine test for each gene product. Again, *T. pallidum* illustrates some of the issues. *T. pallidum* infection is limited to humans so the only true vaccine efficacy test would involve man, but this is not practicable for large-scale screening. Rather, one would like an assay, however much a rough test it is, that allows large-scale screening of gene products for some characteristic of efficacy. This is likely to be in an animal model, not necessarily completely appropriate for man. But as long as it is sufficient to reduce the number of candidate gene products it is useful. Some products that might work as a vaccine will be missed, and others that are found will not pan out in a more rigorous human trial. One must bear in mind that there are likely to be multiple gene products that can function as vaccines and our goal is to find at least some of these, not every one. This smaller number of candidates can then be studied in more detail, with more rigorous approaches, to identify the best ones.

It should also be pointed out that these procedures do not require any knowledge of the function or homologs of the gene products for them to be screened successfully. This brute force science will come up with 'solutions' to important clinical problems without providing 'answers' to the questions of function and mechanism. But clearly when efficacious gene products are identified, they will receive subsequent ardent attention by the research community.

2. The challenges of heterologous expression

Exhaustive screening of the complement of predicted genes of an organism requires expression of each gene so that its product may be tested. As noted above, for many organisms of interest this can only be accomplished in a surrogate host such as *E. coli*. Expression does not mean transcription and translation of the foreign gene in *E. coli* using the native expression signals of the gene. Rather, one attaches *E. coli* expression signals to the coding sequence of the gene, with the aim of producing high quantities of the protein, usually with additional sequences such as a histidine tag to facilitate purification. Expression then refers to the successful detection of the gene product, requiring not only transcription and translation, but also accumulation of the product in a soluble form that can be detected on gels and isolated. While it is easy to construct hybrid genes in which a foreign open reading frame is properly fused to an *E. coli* promoter and translation start sequence, expression of foreign genes in *E. coli* is nevertheless a difficult and unpredictable process.

Our own experiences at expressing *T. pallidum* proteins in *E. coli* are illustrative. Attempts to express five putative hemolysin coding sequences [11] in a selection of vectors gave variable results, some of which are summarized in *table I*. When we initially tried to express these genes in standard high copy number plasmids using the *lac* expression signals, only two of the genes (*tlyC* and *hlyC*) gave good results. In general, expression in vectors using the *lac* promoter and a high copy number replicon was least useful, with instability of the vector a common outcome. This is likely due to the incomplete control of this promoter, resulting in basal synthesis of gene products that may be toxic. Toxicity need not be due to the functional activity of the gene product but could result from aggregation, incomplete secretion, or other detrimental effects of the (inactive) polypeptide on *E. coli* machinery. Regardless, this underscores the importance of regulated expression. We generally had better results with tightly regulated promoters, such as *ara* or T7, and lower copy number vectors. In addition, fusion of additional sequences to the foreign polypeptide could have positive effects, particularly on those products that did not accumulate in the cell despite correct attachment of appropriate expression signals. These gene products may have been lost due to degradation or aggregation in the cell. Thioredoxin fusions showed the best behavior of those that were tried.

Table I. Expression of hemolysin genes in different vectors.

Vector: *Promoter:* *Fusion:*	*pQE-30* Lac *His tag*	*pBAD1HisB* ara *His tag*	*pMAL-p2X* tac *Maltose BP*	*pTYB3* T7 *Intein-CBP*	*pThioHisC* trc *Thioredoxin*
hlyC	+	+	NT	NT	NT
hlyII	–	–	–	+/–	+
hlyA	–	–	–	+	+
hlyB	–	+	+	+	NT
hlyC	+	+	+	+	+

Successful expression is defined as the production of detectable gene product. Some genes may be transcribed and translated properly but the product may be degraded, aggregated, or otherwise undetectable; in these cases expression is scored as negative. The vectors used were: pQE-30 (Qiagen), pBAD1HisB and pThioHisC (Invitrogen), and pMAL-p2X and pTYB3 (New England Biolabs). NT: not tested.

In *figure 2*, the kinetics of expression of three of the hemolysins fused to thioredoxin are shown. The *hlyC* gene (lanes 9–12) expressed well in most vectors. The *hlyA* gene showed a different pattern of expression (lanes 1, 3, 5, 7), with accumulation ceasing soon after induction of expression. This is a pattern suggesting toxicity of the gene product. In contrast the *hlyIII* gene product (lanes 2, 4, 6, 8) was highly aggregated, even under the denaturing conditions used in sample preparation and SDS-polyacrylamide gel electrophoresis. The *hlyIII* product, either alone or fused to other sequences, was not observed after expression in other vectors (*table I*), presumably because it was even more highly aggregated and did not enter the gel. The thioredoxin may reduce aggregation of *hlyIII* sufficiently to allow the polypeptide to enter the gel, albeit still in a partially aggregated form. Fusion of thioredoxin or other sequences to *hlyA* increased production of this polypeptide, although it appeared to be toxic. We speculate that it could be even more toxic in its native state and that the fusions reduced this toxicity, allowing expression to continue for long enough for the protein to be made and accumulate. Another example that is not shown is *hlyB*, which was produced with normal kinetics but at lower levels than *hlyC*. This may indicate some problem with transcription or translation of the coding sequence, or reduced stability of the mRNA or protein. These examples demonstrate that there are a range of problems associated with heterologous expression and it is likely there will not be a single vector solution to expressing all coding sequences from a genome. It should also be noted that the genotype of the host could affect the level of production. Strains with mutations that stabilize mRNA or proteins are of use but are not discussed here.

3. Identifying candidates for immunodiagnostics

Detection methods for diagnostic purposes can be facilitated by a WGS in numerous ways. At a minimum, having the WGS allows one to identify unique sequences that can be used for sensitive PCR-based assays. However, while this readily detects the organism, it provides little other information. Immunodiagnostics, on the other hand, indicate what gene products have been expressed (and recognized by the host) which may be correlated with the stage or type of infection. For a disease such as syphilis, there are characteristic stages, some of which involve long latent periods in which it may be difficult to detect the organism by PCR-based assays. In addition, detecting the organism by PCR does not discriminate between the stages of the infection. However, it is possible that some antigens may be characteristic of each stage and thus immunodiagnostics could not only detect the disease but the stage as well.

Fortunately, expression of active, functional polypeptides are not required for this purpose. Rather, inactive, even incomplete polypeptide sequences will suffice to react with and detect

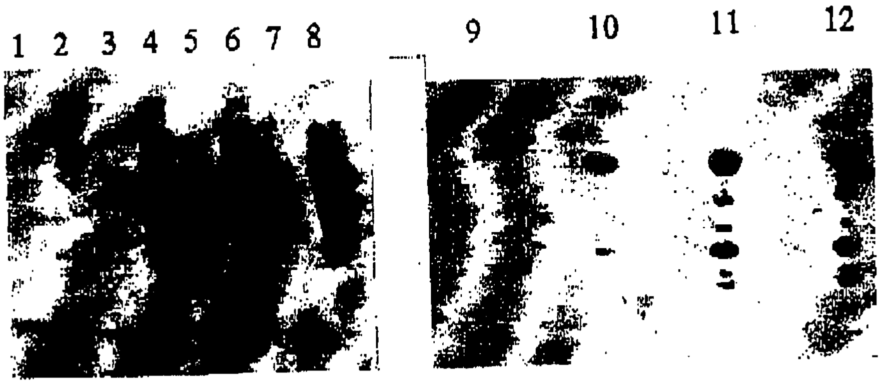


Figure 2. Expression of *hlyA* (lanes 1, 3, 5, 7), *hlyIII* (lanes 2, 4, 6, 8), and *hlyC* (lanes 9, 10, 11, 12) in *E. coli* from the pThioHisC vector at 37°C. Time points were taken every 45 min and subjected to SDS polyacrylamide electrophoresis. Expression products were visualized by western blotting using antibody against thioredoxin. The band in the *hlyA* lanes is the expected size for this fusion protein. The smear of staining in the *hlyIII* lanes is heterogeneous and larger than expected and indicates aggregation of the protein. The upper band in the *hlyC* lanes is the expected size and the lower bands represent degradation products. The two gels were run at different times and the mobilities are not comparable, hence the different relative positions of *hlyA* and *hlyC*.

antibodies in patient sera. The contribution of a WGS to this cause is that it allows the genome to be exhaustively sampled for antigens by expressing each coding sequence as discussed above. Traditionally, antigen-encoding genes are detected by making a shotgun genomic library in a standard vector (e.g. pBR322 or the pUC family) which accomplishes expression in *E. coli* as high copy number vectors using the lac signals. As discussed above, this is not productive for genes that are toxic, or do not accumulate the gene product for various reasons. In addition, this shotgun approach mainly detects the strongest reacting polypeptides. Thus, weak expressors, poor reactors, and difficult to clone genes are selected against, resulting in a significant bias. In contrast, the gene-by-gene approach that is possible with the heterologous expression of all predicted coding sequences from a WGS removes this bias. In the case of *T. pallidum*, for example, there had been several shotgun approaches to antigen detection, but none had detected the hemolysins as antigens [11]. However, using clones expressing each hemolysin and a variety of human syphilitic patient sera, we were able to show reaction of each of these proteins (*table II*). Clearly all of these polypeptides react with at least some of the sera, although only the *tlyC* product showed complete reactivity, with *hlyC* and *hlyIII* products being less completely reactive and *hlyA* and *hlyB* products being least widely reactive. Presumably none of these had been detected in shotgun experiments because of expression difficulties or other issues, such as the strength of reaction. None of the sera used had been staged with respect to the disease but this would be an interesting explanation to account for the more selective reactivity of some of the gene products.

4. Identifying candidates for vaccines

As with tests for immunodiagnostics, determining vaccine candidates does not require expression of active functional polypeptides. Furthermore, the predicted coding sequences can, in principle, be exhaustively sampled using the inactive and/or incomplete polypeptide sequences produced by heterologous expression. The issue here is a suitable in vivo screening method that can suggest efficacy of

Table II. Reactivity of *T. pallidum* hemolysin homologs with sera.

Serum sample	*RPR*	*TlyC*	*HlyC*	*HlyB*	*HlyIII*	*HlyA*
NRS		–	–	–	–	–
IRS			++	–	–	+
1917	NR	++	+	–	++	+
1934	1:32	++	–	–	–	–
1937	1:32	+++	+	–	+	–
1922	NR	++	–	–	+	+
1910	1:2	++	–	–	++	–
1936	1:64	++	–	+	–	–
1926	NR	++	–	–	–	+
1920	1:64	+	+	–	+	–
1902	1:2	+	–	–	+	–
1923	1:64	+	+	–	–	+
1940	1:64	+	–	–	+	–
1912	1:2	+	–	–	+	–
1911	1:2	+++	++	–	+	+
1919	1:256	+	++	–	+	–
1875	1:512	+	+	–	+	+
1947	1:512	+	+	+	–	–
2020	1:512	+++	+++	–	+	–
2068	1:512	+++	++	+	+	+
2054	1:856	++	++	–	+	–
1863	1:2048	++	+	+	+	–

NRS, normal rabbit serum; IRS, immune rabbit serum; all other samples were human syphilitic sera (not staged). RPR, dilution for positive results in the RPR reaction; NR, not reactive. In addition all human sera were positive for the MHA test. The samples of each expressed hemolysin (as a thioredoxin fusion) were run on SDS polyacrylamide gels and subjected to western blotting with each serum sample. The scoring of reactivity from – to +++ was based on qualitative inspection of the blots.

polypeptides and which is suited to large-scale screening. Syphilis is an exclusively human disease, however, and there is no animal model at present. Nevertheless, an example of the direction that may be fruitful is the rabbit skin test shown in *figure 3*. In this case, a different polypeptide, a portion of the *tprF* gene, was expressed as a fusion to a histidine tag, purified and used to immunize a rabbit. The *tprF* gene is a member of a gene family with 12 members in *T. pallidum* [4, 11]. These genes are believed to be important for virulence because of their similarity to the Msp protein of *Treponema denticola*, which has been shown to have several virulence-related interactions with the host [2, 3, 6, 10]. When *T. pallidum* is inoculated on the rabbit back, pustules form but these are reduced in severity in the immunized animal. Although this is a far cry from a model for the multistage syphilis infection in humans, it nevertheless shows that this polypeptide may reduce the pathology associated with initial infection and thus may be useful in preventing initiation or transmission of the infection. While a rabbit is not the ideal experimental animal to use to test the 1 000 coding sequences predicted in the *T. pallidum* WGS, this type of model might be adaptable to smaller animals that are more appropriate for higher throughput analyses.

5. Conclusions and the future: functional analysis

These brief encounters with the *T. pallidum* WGS show the power and influence that this data flood can exert on a field. This organism has been largely impenetrable owing to the fastidious growth requirements. In addition, the novel course of infection and the fact that spirochetes are evolutionarily distant from better studied pathogens (only about half the genes in the WGS match known genes in the databases) prevent precedents from other pathogens

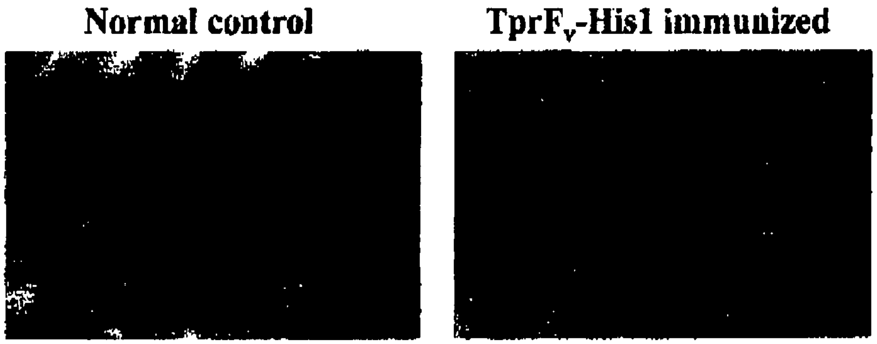


Figure 3. Rabbit challenge study 28 days postinfection. The shaved backs of New Zealand white rabbits, one of which had previously been immunized intradermally with purified TprF protein (expressed and purified as a histidine-tagged fusion), were inoculated with 10^8 treponemes per site. Pustule formation is visible at 28 days in the normal rabbit while the severity of the lesion is reduced in the immunized animal.

being useful as clues to pathogenesis of infection. Yet by focusing on a handful of previously unknown genes, elucidated from the WGS, it is possible to come up with new candidates for immunodiagnostics and vaccines. Clearly, the future of clinical applications from WGSs is a bright one

While this article has focused on the brute force 'solution' side of applications, it is important to consider how WGS information can also be used for functional analysis of the predicted coding sequences. For example, how do we know if any of the hemolysin genes actually code for hemolysins? This requires expression of the coding sequence to produce an intact (or near intact) product that can be tested for activity. As indicated in *table I*, this is a real challenge for some of the genes. The solution to this problem is not clear but there are many avenues to pursue. Perhaps the solution lies with using different vectors, such as those with signal sequences to export proteins, or different hosts, both mutants of *E. coli* as well as other organisms, including bacteria, yeast, or even insect cells. Alternatively, other methods such as phage display technology [7] may obviate the need to produce the protein in a form compatible with the intracellular environment. Or perhaps total in vitro synthesis, using for example a T7 promoter and polymerase and an in vitro translation system, is more appropriate for high-throughput production of gene products. All of these approaches are mainly untested but the desirability of using them to exploit WGSs should stimulate their further development.

Acknowledgments

The authors wish to acknowledge excellent technical support for this work provided by J.K. Howell and the intellectual input from Dr E.J. Sodergren. This work was supported in part by NIH grant AI31068 to CMW.

References

[1] Akerley B.J., Rubin E.J., Camilli A., Lampe D.J., Robertson H.M., Mekalanos J.J., Systematic identification of essential genes by in vitro mariner mutagenesis, Proc. Natl. Acad. Sci. USA 95 (1998) 8927–8932.

[2] Egli C., Leung W.K., Muller K.H., Hancock R.E., McBride B.C., Pore-forming properties of the major 53-kilodalton surface antigen from the outer sheath of *Treponema denticola*, Infect. Immun. 61 (1993) 1694–1699.

[3] Fenno J.C., Muller K.H., McBride B.C., Sequence analysis, expression, and binding activity of recombinant major outer sheath protein (Msp) of *Treponema denticola*, J. Bacteriol. 178 (1996) 2489–2497

[4] Fraser C.M., Norris S.J., Weinstock G.M., White O., Sutton G.G., Dodson R., Gwinn M., Hickey E.K., Clayton R., Ketchum K.A., Sodergren E., Hardham J.M., Mc Leod M.P., Salzberg S., Peterson J., Khalak H., Richardson D., Howell J.K., Chidambaram M., Utterback T., McDonald L., Artiach P., Bowman C., Cotton M.D., Venter J.C., Complete genome sequence of *Treponema pallidum*, the syphilis spirochete. Science 281 (1998) 375–388.

[5] Hensel M., Shea J.E., Gleeson C., Jones M.D., Dalton E., Holden D.W., Simultaneous identification of bacterial virulence genes by negative selection. Science 269 (1995) 400–403.

[6] Mathers D.A., Leung W.K., Fenno J.C., Hong Y., McBride B.C., The major surface protein complex of *Treponema denticola* depolarizes and induces ion channels in HeLa cell membranes, Infect. Immun. 64 (1996) 2904–2910.

[7] Palzkill T., Huang W., Weinstock G.M., Mapping protein-ligand interactions using whole genome phage display libraries. Gene 221 (1998) 79–83.

[8] Schaudinn F., Korrespondenzen. Deut. Med. Wochenschr. 31 (1905) 1728.

[9] Schaudinn F., Hoffman E., Vorlaufiger bericht uber das vorkommen fur spirochaeten in syphilitischen krankheitsprodukten und be papillomen, Arb. Gesundh. Amt. Berlin 22 (1905) 528–534.

[10] Umemoto T., Nakatani Y., Nakamura Y., Namikawa I., Fibronectin binding proteins of a human oral spirochete *Treponema denticola*, Microbiol. Immunol. 37 (1993) 75–78.

[11] Weinstock G.M., Hardham J.M., McLeod M.P., Sodergren E.J., Norris S.J., The genome of *Treponema pallidum*: new light on the agent of syphilis, FEMS Microbiol. Rev. 22 (1998) 323–332.

Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products

Jo Handelsman[1], Michelle R Rondon[1], Sean F Brady[2], Jon Clardy[2] and Robert M Goodman[1]

Cultured soil microorganisms have provided a rich source of natural-product chemistry. Because only a tiny fraction of soil microbes from soil are readily cultured, soil might be the greatest untapped resource for novel chemistry. The concept of cloning the metagenome to access the collective genomes and the biosynthetic machinery of soil microflora is explored here.

Addresses: [1]Department of Plant Pathology, University of Wisconsin–Madison, 1630 Linden Drive, Madison, WI 53706, USA. [2]Department of Chemistry and Chemical Biology, Baker Laboratory, Cornell University, Ithaca, NY 14853, USA.

Chemistry & Biology October 1998, 5:R245–249
http://biomednet.com/elecref/10745521005R0245

Correspondence: Jo Handelsman
E-mail: joh@plantpath.wisc.edu

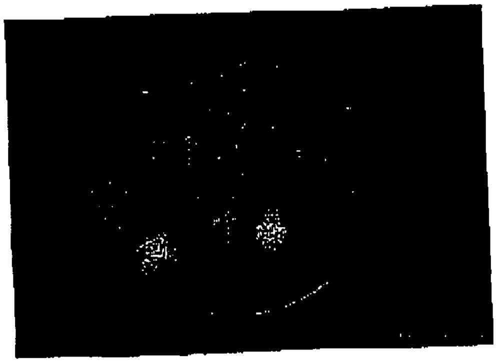
© Current Biology Ltd ISSN 1074-5521

A new frontier of science is emerging that unites biology and chemistry — the exploration of natural products from previously unculturable soil microorganisms. The approach involves directly accessing the genomes of soil organisms that cannot be, or have not been, cultured by isolating their DNA, cloning it into culturable organisms and screening the resultant clones for the production of new chemicals. The excitement surrounding this new field lies in the vast diversity of unknown soil microflora and the chemical richness that they are thought to contain. The methodology has been made possible by advances in molecular biology and eukaryotic genomics, which have laid the groundwork for cloning and functional analysis of the collective genomes of soil microflora, which we term the metagenome of the soil.

Despite the fact that the human species often treats soil like dirt, polluting and degrading it, soil is arguably the most useful and valuable habitat on earth. Humans have used soil for planting crops, for mining for minerals, for building on and for discovering medicinal chemicals for eons. Indeed, cultured soil microorganisms (Figure 1) are the most common source of antibiotics and other medicinal agents of any group of organisms. Pharmaceutical chemists and microbiologists have been culturing the diverse microbes of the soil (Figure 1) and screening them for antibiotic activity since Selman Waksman discovered streptomycin in the actinomycetes (Figure 2) (reviewed in [1]). But, of late, the yield of new natural products from soil microflora has been poor, in part because culturing recovers the same organisms again and again. In actinomycetes, for example, the rediscovery rate for antibiotics is 99% [1].

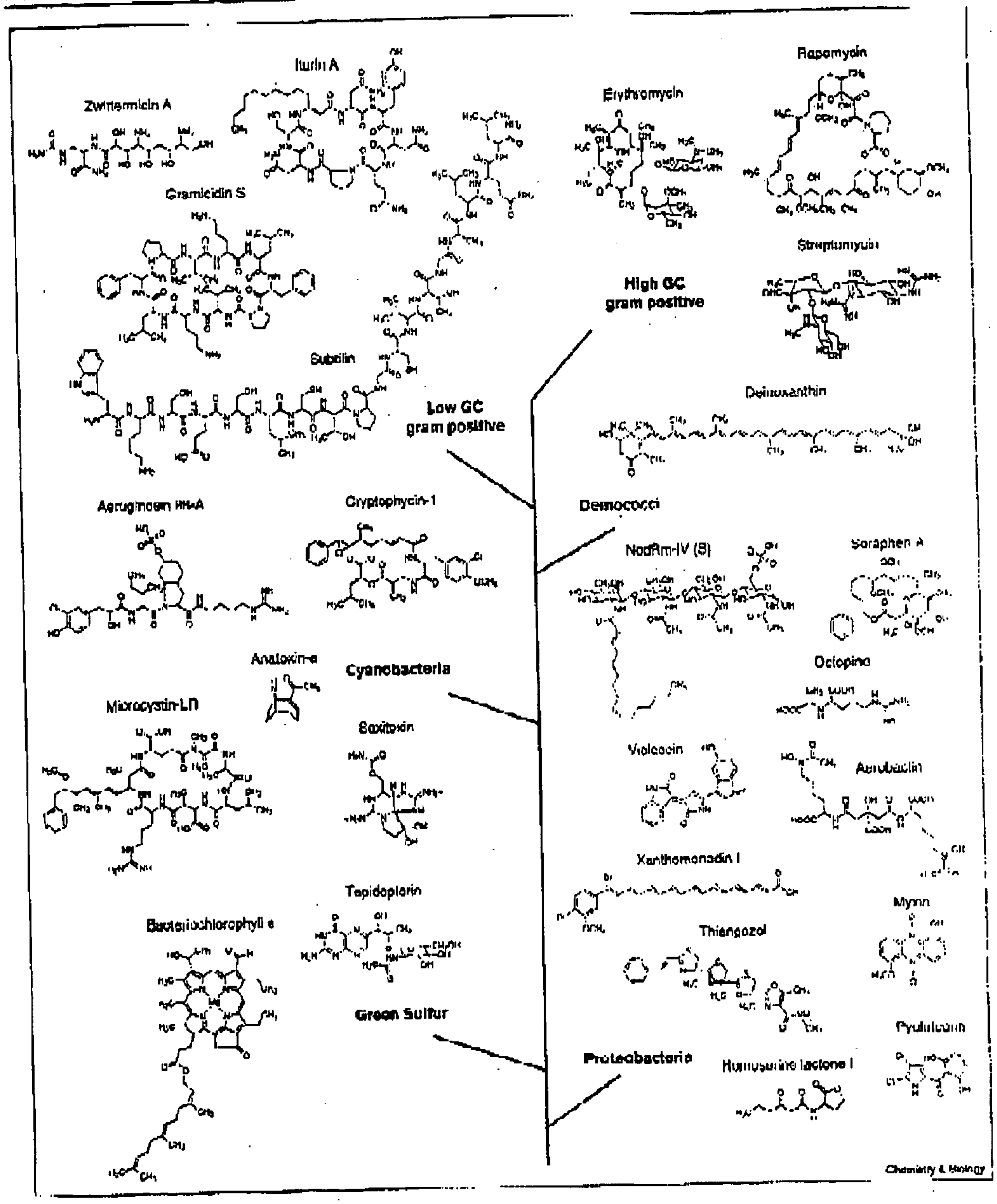
Despite being familiar and useful, soil is also one of the least understood habitats on earth. The last 25 years of research have revealed that culturing is an excellent method to learn a lot about a tiny proportion of the microorganisms on earth [2–7]. Many lines of evidence show that fewer than 0.1% of the microorganisms in soil are readily cultured using current techniques [8–10]. And, most impressively, the other 99.9% of soil microflora is emerging as a world of stunning, novel genetic diversity. New groups of bacteria have been identified in soil that appear to diverge so deeply from the cultured bacteria that they could represent new phyla, or even new kingdoms of life [11–13]. Groups of *Archaea* related to those found thus far only in the open ocean are soil inhabitants around the world [14,15]. Estimates are that a gram of soil might contain 1,000–10,000 species of unknown prokaryotes [8]. There is likely to be further diversity within species, which current phylogenetic analysis cannot resolve. Because microbes, generally, have great genetic diversity — soil carries the highest populations of microbes of any habitat [16] — and microbes cultured from soil have revealed tremendous chemical virtuosity and utility, the vast majority of as yet unknown microbes could well be a far greater source of new molecular structures than any habitat on earth. Tapping into this source should be a great, joint adventure for biologists and chemists.

Figure 1



Morphological diversity typical of microorganisms cultured from soil on a broad spectrum medium, tryptic soy agar.

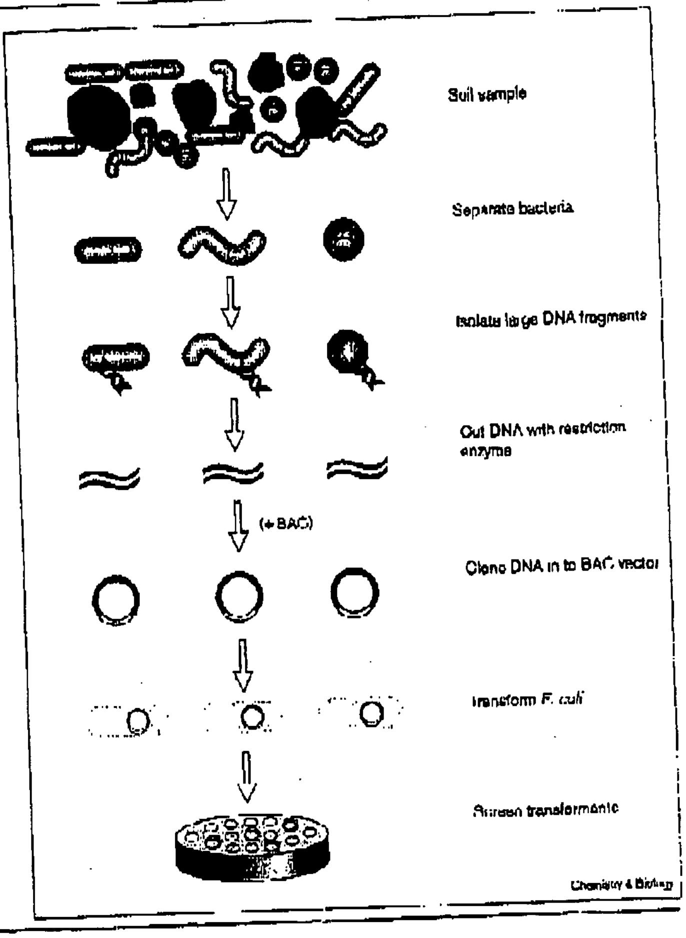
Figure 2


Zwittermicin A
Iturin A
Erythromycin
Rapamycin
Gramicidin S
Streptomycin
High GC gram positive
Subtilin
Low GC gram positive
Deinoxanthin
Aeruginosin 98-A
Cryptophycin-1
Deinococci
NodRm-IV (S)
Soraphen A
Anatoxin-a
Cyanobacteria
Octopine
Microcystin-LR
Saxitoxin
Violacein
Agrobactin
Xanthomonadin I
Myxin
Tepidopterin
Bacteriochlorophyll e
Thiangazol
Green Sulfur
Proteobacteria
Homoserine lactone I
Pyoluteorin
Chemistry & Biology

Examples of the chemical diversity in some of the major phyla of bacteria.

Figure 3

Cloning the metagenome is our process for isolating new pathways for the synthesis of bioactive molecules from noncultured soil microorganisms. DNA is extracted directly from soil, using gentle methods to preserve high-molecular-weight DNA. The DNA is cut using a restriction enzyme and cloned into a bacterial artificial chromosome (BAC), a vector which can carry large fragments of DNA in *E. coli*. The BAC clones are then screened for biological activity and for the production of novel natural products.


Soil sample
Separate bacteria
Isolate large DNA fragments
Cut DNA with restriction enzyme
(+BAC)
Clone DNA in to BAC vector
Transform *E. coli*
Screen transformants

Chemistry & Biology

Accessing the chemistry of microbial diversity presents an enticing but difficult challenge, in part because most of the novel structures are likely to be in organisms present in low abundance in the soil. Developing methods to culture the enormous diversity of soil microflora will be slow and tedious and will require more knowledge of the physiology of the unknown microbes than is presently in hand. What is needed is a more direct, global and rapid method to access the genetic riches of soil microflora.

We have embarked on an effort to access the chemical diversity of soil life by cloning the metagenome of the soil without first culturing the microflora, treating the metagenome as a genomic unit. The strategy is to isolate metagenomic DNA directly from soil, clone it in large pieces into a readily cultured organism such as *Escherichia coli*, and screen the clones for biological activity (Figure 3). The first hurdle is to clone and maintain large pieces of DNA. The present-day vectors of choice for such endeavours are the bacterial artificial chromosome (BAC) vectors, which were originally developed for cloning eukaryotic genome fragments. BACs are maintained at low copy number in *E. coli* and can carry DNA inserts as large as 350 kilobases [17]. Although used extensively for animal and plant genomics, BACs have not been applied much to bacterial genomics. The metagenome of the soil presents challenges of size and complexity similar to those associated with eukaryotic genomics, so the BAC vectors are appropriate tools for the task. The technical challenge inherent in this approach is maintaining the large size of the DNA fragments while removing non-DNA soil material that inhibits cloning. DNA fragments of up to 40 kilobases have been cloned directly from aquatic environments [18] and we recently cloned fragments greater than 70 kilobases directly from soil (our unpublished observations).

Expression of anonymous microbe genes in the host cell is required in order for the production and detection of new chemicals to be possible. Although many genes will not be expressed in any given host, such as *E. coli*, many others will be. Our own data show that the diversity of uncultured soil microorganisms in the phylum of Proteobacteria, which contains *E. coli*, is surprisingly high, suggesting that, even if expression of genes was obtained from only Proteobacteria, the clone bank would provide access to fantastic genetic diversity. The known Proteobacteria are a diverse group of prokaryotes including many that produce interesting natural products (Figure 2); the myxobacteria are an example of a group within the Proteobacteria that was only recently recognized to produce diverse and valuable natural products [19].

There is strong evidence that gene expression in *E. coli* will not be limited to genes from the Proteobacteria. Genes of diverse prokaryotes, from *Thermus* to *Corynebacterium*, can be expressed in *E. coli* by simply introducing the relevant genes — no special tinkering or engineering of the DNA is required to obtain expression [20–27]. This suggests that gene expression will not be a major barrier to obtaining functional clones even in *E. coli*. For example, we constructed a BAC library in *E. coli* with DNA from *Bacillus cereus*, a Gram-positive bacterium that is phylogenetically quite distant from *E. coli* (the distance between them is equivalent to the distance between humans and paramecia). In screens for *B. cereus* traits in the library of *B. cereus* DNA in the *E. coli* host, we found that more than half of the traits tested were expressed in the library, some at quite high levels (M.R.R., S.J. Raffel, R.M.G. and J.H, unpublished observations). We believe, therefore, that this is a promising approach for cloning and expressing of genes from diverse organisms. Moreover, the spectrum of gene expression might be broadened by constructing additional libraries in *Streptomyces*, *Bacillus* and *Archaea*.

Some features of known biosynthetic pathways of secondary metabolites from bacteria make the proposed approach feasible. First, the genes for natural-product biosynthetic pathways are usually clustered in prokaryotes [28–36], making it possible to clone an entire pathway into a BAC vector on a contiguous piece of DNA. Second, for natural products that are potentially toxic to prokaryotes, such as antibiotics, the biosynthetic clusters are linked to genes for resistance to the natural product, so that the organism carrying the biosynthetic machinery does not die because of inhibition by the expressed product [29,30]. It is reasonable, therefore, to expect that if a pathway for a natural product were expressed in *E. coli*, the resistance mechanism would be as well, thereby protecting the host cell.

Recent advances in screening for biological activity make cloning the metagenome and screening the resultant clones for natural products both timely and practical. High-throughput screening makes it feasible to test the 1,000,000 clones that are likely to be required to cover the metagenome of the soil. The sensitivity of modern assays for biological activity, particularly those assays conducted on a nanoscale, provides a means for identifying clones that produce or export tiny amounts of an active molecule — only moderate expression of heterologous genes in the host bacterium is therefore required [37–39]. The use of *E. coli* as the host cell extends the power of this approach, given that *E. coli* is commonly used in industrial fermentation, so sophisticated methods that facilitate batch production, separations, as well as downstream processing are well established. This means that many of the development stages for commercial production of useful products have already been carried out before the genes are cloned, offering an advantage over natural products derived directly from 'wild' organisms that might be difficult to tame for industrial purposes. The methods developed for the discovery of new natural product synthesis pathways from soil microorganisms can, in the future, be applied to other habitats, such as the microflora of insects or marine animals, which are thought to be a good source of novel compounds but are often difficult to culture [40].

Will the genetic diversity contained in the soil metagenome reveal a new level of chemical diversity in the encoded natural products? Experience suggests that it will. For example, marine organisms began to be intensely examined for natural products roughly 25 years ago, and the result was the identification of an impressive number of dramatically new compounds in a remarkably short time [41]. Although the question of the chemical diversity of the soil metagenome might still be open, we shouldn't have to wait long for an answer.

The enormous potential of soil microbial resources can only be tapped through the combined efforts of chemists and biologists. Both chemistry and biology have powerful and innovative techniques to bring to bear on the problem, and the cross-fertilization provided by jointly exploring the soil metagenome will, we hope, drive even greater innovation. Of course, such alliances will require new mechanisms of funding along with appropriate training to nurture collaborations across traditional disciplinary divides. But the likely discoveries — both in fundamental knowledge and in terms of therapeutically useful molecules — call for the rapid formation of such alliances, which should provide exhilarating experiences for all involved.

Acknowledgements
This work was supported by Ariad Pharmaceuticals and the National Institutes of Health.

References

1. Zaehner, H. & Fiedler, F.P. (1990). Fifty years of antimicrobials: past perspectives and future trends. In *The Need for New Antibiotics: Possible Ways Forward.* Fifty-Third Symposium of the Society for General Microbiology, Bath, England, UK. (Hunter, P.A., Darby, G.K. & Russell, N.J., eds.), pp. 67-84, Cambridge University Press, Cambridge, UK.
2. Amann, R.I., Ludwig, W. & Schleifer, K.-H. (1995). Phylogenetic identification and *in situ* detection of individual microbial cells without culturing. *Microbiol. Rev.* 59, 143-169.
3. DeLong, E.F. (1992). *Archaea* in coastal marine environments. *Proc. Natl Acad. Sci. USA* 89, 5685-5689.
4. Giovannoni, S.J., Britschgi, T.B., Moyer, C.L. & Field, K.G. (1990). Genetic diversity in Sargasso Sea bacterioplankton. *Nature* 345, 60-63.
5. Pace, N.R. (1996). New perspective on the natural microbial world: molecular microbial ecology. *ASM News* 62, 463-470.
6. Stahl, D.A., Lane, D.J., Olsen, G.J. & Pace, N.R. (1985). Characterization of a Yellowstone hot spring microbial community by 5S rRNA sequences. *Appl. Environ. Microbiol.* 49, 1379-1384.
7. Ward, D.M., Weller, R. & Bateson, M.M. (1990). 16S rRNA sequences reveal numerous uncultured microorganisms in a natural community. *Nature* 345, 63-65.
8. Torsvik, V., Goksøyr, J. & Daae, F.L. (1990). High diversity in DNA of soil bacteria. *Appl. Environ. Microbiol.* 56, 782-787.
9. Torsvik, V., Goksøyr, J., Daae, F.L., Sørheim, R., Michalsen, J. & Salte, K. (1994). Use of DNA analysis to determine the diversity of microbial communities. In *Beyond the Biomass.* (Ritz, K., Dighton, J. & Giller, K.E., eds), pp. 39-48, John Wiley & Sons, Chichester.
10. Torsvik, V., Sørheim, R. & Goksøyr, J. (1996). Total bacterial diversity in soil and sediment communities-a review. *J. Industr. Microbiol.* 17, 170-178.
11. Borneman, J., *et al.*, & Triplett, E.W. (1996). Molecular microbial diversity of an agricultural soil in Wisconsin. *Appl. Environ. Microbiol.* 62, 1935-1943.
12. Liesack, W. & Stackebrandt, E. (1992). Occurrence of novel groups of the domain Bacteria as revealed by analysis of genetic material isolated from an Australian terrestrial environment. *J. Bacteriol.* 174, 5072-5078.
13. Kuske, C.R., Barns, S.M. & Busch, J.D. (1997). Diverse uncultured bacterial groups from soils of the arid southwestern United States that are present in many geographic regions. *Appl. Environ. Microbiol.* 63, 3614-3621.
14. Bintrim, S.B., Donohue, T.J., Handelsman, J., Roberts, G.P. & Goodman, R.M. (1997). Molecular phylogeny of *Archaea* from soil. *Proc. Natl Acad. Sci. USA* 94, 277-282.
15. Ueda, T., Suga, Y. & Matsuguchi, T. (1995). Molecular phylogenetic analysis of a soil microbial community in a soybean field. *Eur. J. Soil Sci.* 46, 415-421.
16. Whitman, W.B., Coleman, D.C. & Wiebe, W.J. (1998). Prokaryotes: the unseen majority. *Proc. Natl Acad. Sci. USA* 95, 6578-6583.
17. Shizuya, H. & Simon, M. (1992). Cloning and stable maintenance of 300-kilobase-pair fragments of human DNA in *Escherichia coli* using an F-factor-based vector. *Proc. Natl. Acad. Sci. USA* 89, 8794-8797.
18. Stein, J.L., Marsh, T.L., Wu, K.Y., Shizuya, H. & DeLong, E.F. (1996). Characterization of uncultured prokaryotes: isolation and analysis of a 40-kilobase-pair genome fragment from a planktonic marine archaeon. *J. Bacteriol.* 178, 591-599.
19. Reichenbach, H. & Hofle, G. (1993). New secondary metabolites from Myxobacteria. In *Third International Conference on the Biotechnology of Microbial Products: Novel Pharmacological and Agrobiological Activities.* Developments in Industrial Microbiology Series Volume 33. (Gullo, V.P., Hunter-Cevera, J.C., Cooper, R. & Johnson, R.K., eds.), Society for Industrial Microbiology, Fairfax, VA.
20. Black, C., Fyfe, J.A.M. & Davies, J.K. (1995). A promoter associated with the neisserial repeat can be used to transcribe the uvrB gene from *Neisseria gonorrhoeae*. *J. Bacteriol.* 177, 1952-1958.
21. Chávez, S., Reyes, J.C., Chauvat, F., Florencio, F.J. & Candau, P. (1995). The NADP glutamate dehydrogenase of the cyanobacterium *Synechocystis* 6803: cloning, transcriptional analysis and disruption of the gdhA gene. *Plant Mol. Biol.* 28, 173-188.
22. Ding, M. & Yelton, D.B. (1993). Cloning and analysis of the leuB gene of *Leptospira interrogans* serovar *pomona*. *J. Gen. Microbiol.* 139, 1093-1103.
23. Ferreyra, R.G., Soncini, F.C. & Viale, A.M. (1993). Cloning, characterization, and functional expression in *Escherichia coli* of chaperonin (groESL) genes from the phototrophic sulfur bacterium *Chromatium vinosum*. *J. Bacteriol.* 175, 1514-1523.
24. Pátek, M., Eikmanns, B.J., Pátek, J. & Sahm, H. (1996). Promoters from *Corynebacterium glutamicum*: cloning, molecular analysis, and search for a consensus motif. *Microbiology* 142, 1297-1309.
25. Sarker, M.R., Akimoto, S., Ono, T., Kinouchi, T. & Ohnishi, Y. (1995). Molecular cloning of the leuB gene from *Bacteroides fragilis* by functional complementation in *Escherichia coli*. *Microbiol. Immunol.* 39, 19-26.
26. Söhling, B. & Gottschalk, G. (1996). Molecular analysis of the anaerobic succinate degradation pathway in *Clostridium kluyveri*. *J. Bacteriol.* 178, 871-880.
27. Sterner, R., Dahm, A., Darimont, B., Ivens, A., Liebl, W. & Kirschner, K. (1995). βα-barrel proteins of tryptophan biosynthesis in the hyperthermophile *Thermotoga maritima*. *EMBO J.* 14, 4395-4402.
28. Aharonowitz, Y., Cohen, G. & Martin, J.F. (1992). Penicillin and cephalosporin biosynthetic genes: structure, organization, regulation and evolution. *Annu. Rev. Microbiol.* 46, 461-495.
29. Chater, K.F. & Bruton, C.J. (1985). Resistance, regulatory and production genes for the antibiotic methylenomycin are clustered. *EMBO J.* 4, 1893-1897.
30. Distler, J., Braun, C., Ebert, A., & Piepersberg, W. (1987). Gene cluster for streptomycin biosynthesis in *Streptomyces griseus*: analysis of a central region including the major resistance gene. *Mol. Gen. Genet.* 208, 204-210.
31. Donadio, S., Staver, M.J., McAlpine, J.B., Swanson, S.J., & Katz, L. (1991). Modular organization of genes required for complex polyketide biosynthesis. *Science* 252, 675-679.
32. Lomovskaya, N., *et al.*, & Hutchinson, C.R. (1998). The *Streptomyces peucetius* dpsY and dnrX genes govern early and late steps of daunorubicin and doxorubicin biosynthesis. *J. Bacteriol.* 180, 2379-2386.
33. Martin, J.F. & Liras, P. (1989). Organization and expression of genes involved in the biosynthesis of antibiotics and other secondary metabolites. *Annu. Rev. Microbiol.* 43, 173-206.
34. Martin, J.F. (1992). Clusters of genes for the biosynthesis of antibiotics: regulatory genes and overproduction of pharmaceuticals. *J. Industr. Microbiol.* 9, 73-90.
35. Skaugen, M., Abildgaard, C.I.M. & Nes, I.F. (1997). Organization and expression of a gene cluster involved in the biosynthesis of the lantibiotic lactocin S. *Mol. Gen. Genet.* 253, 674-686.
36. Krause, M. & Marahiel, M.A. (1988). Organization of the biosynthesis genes for the peptide antibiotic gramicidin S. *J. Bacteriol.* 170, 4669-4674.
37. Borchardt, A., Liberles, S.D., Biggar, S.R., Crabtree, G.R. & Schreiber, S.L. (1997). Small molecule-dependent genetic selection in stochastic nanodroplets as a means of detecting protein-ligand interactions on a large scale. *Chem. Biol.* 4, 961-968.
38. Huang, J. & Schreiber, S.L. (1997). A yeast genetic system for selecting small molecule inhibitors of protein-protein interactions in nanodroplets. *Proc. Natl Acad. Sci. USA* 94, 13396-13401.
39. You, A.J., Jackman, R.J., Whitesides, G.M. & Schreiber, S.L. (1997). A miniaturized arrayed assay format for detecting small-molecule–protein interactions in cells. *Chem. Biol.* 4, 969-975.
40. Ohkuma, M. & Kudo, T. (1996). Phylogenetic diversity of the intestinal bacterial community in the termite Reticulitermes speratus. *Appl. Environ. Microbiol.* 62, 461-468.
41. Newman, D.J., Jensen, P.R., Clement, J.J. & Acebal, C. (1989). Novel activities from marine derived microorganisms. In *Novel Microbial Products for Medicine and Agriculture.* (Demain, A.L., Somkuti, G.A., Hunter-Cevera, J.C. & Rossmoore, H.W., eds.), pp. 239-252, Elsevier, Amsterdam.

Proc. Natl. Acad. Sci. USA
Vol. 96, pp. 6451–6455, May 1999
Microbiology

Toward functional genomics in bacteria: Analysis of gene expression in *Escherichia coli* from a bacterial artificial chromosome library of *Bacillus cereus*

MICHELLE R. RONDON, SANDRA J. RAFFEL, ROBERT M. GOODMAN*, AND JO HANDELSMAN

Department of Plant Pathology, University of Wisconsin-Madison, 1630 Linden Drive, Russell Laboratories, Madison, WI 53706

Communicated by Donald R. Helinski, University of California at San Diego, La Jolla, CA, March 30, 1999 (received for review November 10, 1998)

ABSTRACT As the study of microbes moves into the era of functional genomics, there is an increasing need for molecular tools for analysis of a wide diversity of microorganisms. Currently, biological study of many prokaryotes of agricultural, medical, and fundamental scientific interest is limited by the lack of adequate genetic tools. We report the application of the bacterial artificial chromosome (BAC) vector to prokaryotic biology as a powerful approach to address this need. We constructed a BAC library in *Escherichia coli* from genomic DNA of the Gram-positive bacterium *Bacillus cereus*. This library provides 5.75-fold coverage of the *B. cereus* genome, with an average insert size of 98 kb. To determine the extent of heterologous expression of *B. cereus* genes in the library, we screened it for expression of several *B. cereus* activities in the *E. coli* host. Clones expressing 6 of 10 activities tested were identified in the library, namely, ampicillin resistance, zwittermicin A resistance, esculin hydrolysis, hemolysis, orange pigment production, and lecithinase activity. We analyzed selected BAC clones genetically to identify rapidly specific *B. cereus* loci. These results suggest that BAC libraries will provide a powerful approach for studying gene expression from diverse prokaryotes.

The massive accumulation of prokaryotic DNA sequences, including an increasing number of complete genome sequences, is revolutionizing the practice and potential of microbiology. Classical genetic techniques, such as those developed for *Escherichia coli*, *Salmonella typhimurium*, and *Bacillus subtilis*, are no longer the only powerful methods for the investigation of gene expression and function in bacterial systems. Because tools for genetic analysis of many prokaryotes are lacking, we are interested in the development of broadly applicable systems for the investigation of the biology of diverse prokaryotes. A promising technology emerging from genomics is the ability to investigate rapidly the biological features of any organism of choice, without prior development of a specific genetic system for that organism. Here we describe the application of a powerful tool in genomic technology, the bacterial artificial chromosome (BAC), to the study of *Bacillus cereus*, a bacterium for which classical genetic tools are not available.

The BAC vector, based on the *E. coli* F factor, was developed for cloning large fragments of eukaryotic DNA in *E. coli* (1). BACs replicate at a copy number of one to two per cell and are maintained very stably in the cell. BAC plasmids can be isolated easily from chromosomal DNA and purified in sufficient quantity for analysis and sequencing. These properties separate BACs from other cloning vectors such as yeast artificial chromosomes and cosmids, which also may suffer from instability and chimera problems (1, 2).

BAC libraries of genomic DNA from numerous plant, animal, and fungal species have been constructed and are becoming the preferred approach in many large-scale sequencing projects (3–5). Molecular techniques developed with BAC technology include methods for introduction of reporter genes into mammalian systems, *in vivo* complementation of mutations, and *in vivo* and *in vitro* "retrofitting" protocols to add new sequence elements to BAC plasmids (6–9).

To date, BAC technology has been applied in a limited way to prokaryote genomics (10, 11). We believe it deserves much wider appreciation, because it offers significant advantages for cloning and analysis of prokaryotic genomes. BACs can be used to clone complex loci, such as biosynthetic pathways, secretion systems, or pathogenicity islands, because the average insert size of a BAC clone is usually greater than 100 kb and because the genes for many bacterial pathways are clustered in the genome. Because BAC inserts are large, a relatively small number of clones is required to provide complete coverage of a bacterial genome, minimizing the amount of work required to screen a BAC library for complex functions or to construct a minimum overlap library. Most significantly, and in contrast to eukaryotic BAC libraries, gene expression from bacterial BAC libraries can be detected in the host strain, thus providing a surrogate system for the analysis of complex pathways from poorly studied, difficult to manipulate, or even uncultured prokaryotes from environmental samples. Therefore, bacterial BAC libraries can serve to archive DNA for genomics purposes and, concurrently, can be used to analyze gene expression, a first step in functional genomics analysis.

To test the applicability of BACs to the study of bacterial functional genomics, we constructed a BAC library in *E. coli* from genomic DNA of the Gram-positive bacterium *B. cereus*. We screened the library for characteristic *B. cereus* activities expressed in *E. coli* and report that a significant number of such activities can be detected in a small library. Our results suggest that BACs provide a useful technique for heterologous expression and functional genomics in prokaryotes.

MATERIALS AND METHODS

Bacterial Strains and Plasmids Used. *B. cereus* strain UW85 was described previously (12). *E. coli* strain DH10B, the host strain for the BAC library, and the BAC vector pBeloBAC11 were obtained from H. Shizuya (1, 3).

Preparation of High Molecular Weight DNA from B. cereus. High molecular weight DNA was prepared by a modification of the method described in ref. 13. *B. cereus* cells were grown for 5 h at 28°C in LB to a density of 3×10^9 cells per ml. Cells were harvested by centrifugation and washed once in 1 vol of

---

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

PNAS is available online at www.pnas.org.

Abbreviation: BAC, bacterial artificial chromosome.

*To whom reprint requests should be addressed. e-mail: RGoodman@facstaff.wisc.edu.

buffer A (50 mM Tris·HCl, pH 8.0/1 M NaCl). A sample was taken to determine cell density, and the remaining culture was stored as a cell pellet at −20°C. The pellet was resuspended to a final concentration of 2 × 10^9 cells per ml and mixed with an equal volume of molten 1.6% SeaPlaque GTG agarose (FMC), pipetted into plug molds, and allowed to cool. The resulting plugs were incubated for 24 h at 37°C in 2 vol of buffer B [50 mM Tris·HCl, pH 8.0/100 mM EDTA, pH 8.0/100 mM NaCl/0.2% sodium deoxycholate/0.5% 20 cetyl ether (Brij-58)/0.5% *S*-laurylsarcosine] with 5 mg/ml lysozyme added. The plugs were transferred to 2 vol of buffer C (50 mM Tris·HCl, pH 8.0/500 mM EDTA, pH 8.0/100 mM NaCl/0.5% *S*-laurylsarcosine/0.2 mg/ml proteinase K) and incubated for 24 h at 50°C. This step was repeated once. Plugs were washed extensively with TE (10 mM Tris/1 mM EDTA, pH 8.0), followed by inactivation of proteinase K with PMSF. Plugs were stored at 4°C in 10 mM Tris·HCl, pH 8.0/50 mM EDTA, pH 8.0.

Digestion of *B. cereus* DNA *in Situ* and Isolation of Sized Fragments. Partial *Hin*dIII digestion was used to prepare large fragments of DNA from the plugs. Agarose plugs first were incubated in two changes of 1 ml of TE per plug for 5 h at room temperature to remove storage buffer components. Partial digestion by limiting Mg^{2+} concentration was performed as described in ref. 14. The plugs were loaded onto a 1% SeaPlaque agarose gel, and the DNA was size-fractionated by pulsed-field gel electrophoresis. Gel slices containing DNA of the appropriate size were cut out and digested with GELase (Epicentre Technologies, Madison, WI) before ligation. Separation conditions were varied to optimize removal of DNA fragments smaller than 100 kb.

Preparation of BAC Vector, Ligation, and Transformation. Protocols for library construction were taken from the URL http://www.tree.caltech.edu with the following modifications. After the plasmid pBeloBac11 was purified with the Qiagen Plasmid Maxi Kit, it was purified further by LiCl precipitation, RNase treatment, and polyethylene glycol precipitation, as detailed in ref. 15. Finally, the plasmid was treated with PlasmidSafe DNase (Epicentre Technologies) as recommended by the manufacturer. Plasmid DNA (10 μg) was digested with *Hin*dIII and dephosphorylated with HK Phosphatase (Epicentre Technologies), followed by phenol/chloroform extraction and ethanol precipitation. Ligations and transformations were performed as described in the above mentioned URL. One microliter of ligation mix was used to transform 50 μl of DH10B competent cells by electroporation with a Bio-Rad GenePulser instrument. Samples of 100 μl were spread on LB plates containing 12.5 μg/ml chloramphenicol, 25 μg/ml isopropyl β-D-thiogalactopyranoside, and 50 μg/ml 5-bromo-4-chloro-3-indolyl β-D-galactopyranoside. After 36 h at 37°C, white colonies were picked for further analysis.

Plasmid Preparation of BAC DNA. DNA was isolated as described in ref. 15, with modifications. Cells from an overnight culture (1.5 ml) were centrifuged in a 1.5-ml microcentrifuge tube and resuspended in 100 μl of resuspension solution (50 mM glucose/10 mM EDTA/10 mM Tris·Cl, pH 8.0). Freshly prepared 0.2 M NaOH/1% SDS (200 μl) was added and the suspension was mixed by inversion of the tube. After a 5-min incubation at room temperature, 150 μl of 7.5 M ammonium acetate and 150 μl of chloroform were added and mixed by inversion of the tube. The samples were incubated for 10 min on ice and then centrifuged at 14,000 rpm for 10 min in an Eppendorf model 5415C microcentrifuge. The supernatant fluid was added to 200 μl of 30% polyethylene glycol 8000/1.5 M NaCl, mixed by inversion, and incubated on ice for 15 min. Precipitated BAC DNA was collected by centrifugation for 10 min, all of the remaining liquid was removed from the tube, and the pellet was resuspended in 25 μl of sterile water. One-fifth of the sample was used per restriction digest. Digests were analyzed by pulse-field gel electrophoresis on a Pharmacia GeneNavigator using hexagonal electrodes, with the following parameters: 1% SeaKem agarose gel (FMC); linear pulse time ramp from 5 to 15 s; 18-h run time; 165 V.

Activity Screens. Media for testing esculin hydrolysis, lecithinase activity, starch hydrolysis, casein hydrolysis, and lipase activity were as described (16). Ampicillin resistance was tested on LB plates containing 50 μg/ml sodium ampicillin. Hemolytic activity was determined on sheep/veal blood agar plates prepared by the Wisconsin State Hygiene Laboratory. *zmaR*-containing clones were identified by colony hybridization and confirmed to be zwittermicin A-resistant by radial streak assay (17).

λ-Tn*phoA* Mutagenesis. Mutagenesis was performed as described (18, 19). Mid-logarithmic-phase cells were infected with λ-Tn*phoA* at a ratio of 1:1, incubated for 2 h at 28°C, and then plated on LB plates with chloramphenicol (Cm, 12.5 μg/ml) and kanamycin (Km, 50 μg/ml). Colonies were pooled and plasmid DNA was prepared from the pools as described above. One microliter of plasmid DNA was used for transformation of DH10B by electroporation. Cm^R Km^R colonies were selected and then replicated to indicator medium to screen for the loss of function phenotype.

Mini-Tn*10*-Kan Mutagenesis. We transformed pLOFKm (20) by electroporation into competent cells of strains containing the BAC of interest. Cm^R Km^R transformants were selected, pooled, and processed as described above for λ-Tn*phoA* mutants.

Sequencing. BAC DNA for sequencing was prepared by using the Qiagen Plasmid Midi kit, following the protocol for BACs as specified by the manufacturer. The final pellet was suspended in 1 ml of sterile water, precipitated with 2 vol of ethanol, and resuspended in 75 μl of 10 mM Tris·Cl, pH 8.0. Sequencing reactions were performed with 2 μg of DNA and 10 pmol of primer, using BigDye reaction mix (Perkin–Elmer). Reaction products were purified with MicroSpinPreps (Pharmacia) or with CentriSep columns (Princeton Separations, Adelphia, NJ). Sequencing reactions were run on an Applied Biosystems 377 sequencer at the University of Wisconsin Biotechnology Center. Standard T7 and SP6 primers (Promega) were used to generate end sequence, and *phoA* primers were *phoA1* (5′-AATATCGCCCTGAGCAGCCCG-3′) (21) and *phoA4* (5′-TAGGAGGTCACATGGAAGTCAGATC-3′) (22).

Clone Grouping. Denatured miniprep DNA (2 μl) of each BAC clone was spotted to a nylon membrane (Magnagraph; Micron Separations) previously wet with water and then with 2× SSC (1.5 M sodium chloride/30 mM sodium citrate, pH 7). Membranes were kept wet while spotting the DNA by placing them on filter paper soaked with 2× SSC. DNA was crosslinked to the membrane by using a Stratalinker. The 700- and 900-kb fragments were generated by *Not*I digestion of *B. cereus* chromosomal DNA and separated by pulse field gel electrophoresis using 0.5× TBE buffer modified to contain only 0.1× EDTA. Run parameters were 11°C for 13 h at 70-s pulse time and then 11 h at 120-s pulse time. The gel was stained briefly with ethidium bromide, and gel slices were cut out. After dialysis in sterile, deionized water to remove excess ethidium bromide, the DNA was labeled *in situ* by using the Genius Kit (Boehringer Mannheim), according to the protocol for in-gel labeling from the FMC catalog. Hybridizations were done according to the protocol for Magnagraph membranes.

RESULTS

Construction of a *B. cereus* BAC Library. The BAC library consists of 323 clones, containing approximately 30 Mb of *B. cereus* DNA. We estimated the size of the UW85 genome at 5.5 Mb, based on *Not*I and *Sfi*I digests (data not shown); on this basis, our library represents a 5.75-fold coverage of the UW85

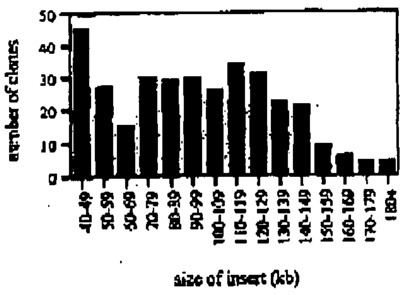


FIG. 1. Size distribution of BAC clones. Insert size was estimated by digestion of the plasmids with *Not*I, followed by analysis of the digestion products by pulse-field gel electrophoresis.

genome. The range of insert sizes is from 40 to greater than 175 kb (Fig. 1). The average insert size was 98 kb. Greater than half (216 clones) of the BAC clones contain insert DNA greater than 80 kb in size. We estimate that the probability of the library containing any particular 1-kb gene is 99.7% (23).

Numerous molecular techniques, such as contig building, end-sequence analysis, hybridization, and clone pooling, have been applied successfully to BAC libraries for rapid identification of a clone of interest, physical mapping, and comparative genomics (24, 25). We identified a set of BAC clones all hybridizing to a single genomic *Not*I fragment from *B. cereus*. Using a 700-kb *Not*I fragment as a probe, we identified 37 clones that hybridized to this fragment (Fig. 2). These clones contain approximately 3,060 kb of DNA, representing 4.4-fold coverage of the probe fragment. The fragment represents 12.7% of the genome, and the clones identified contain 10.4% of the DNA in the library, indicating that the genomic DNA in this fragment is represented proportionately in the library. We also probed by using a 900-kb fragment and found 19 clones that hybridized to this fragment (data not shown). This represents a 2.7-fold coverage of the fragment. These clones contain 8% of the DNA in the library, although the fragment represents 16% of the genome. This section of the genome therefore may be underrepresented in the library.

Potential of BACs as Surrogate Expression Vectors for Analysis of Prokaryotic Genes. We estimated the frequency of gene expression in the library by testing it for easily detectable *B. cereus* activities. The BAC library was replicated by using 48-prong replicators to various media to test for the expression of *B. cereus* activities in *E. coli*. Of nine activities tested by the use of specific indicator media, six were found in the library (Table 1). Additionally, zwittermicin A-resistant clones were identified by hybridization to a *zmaR*-containing probe and confirmed to be resistant by radial streak assay (E. A. Stohl and J.H., unpublished results). In all cases the BAC clones were isolated from putative positive colonies and retransformed into DH10B, and the activity was found to be BAC-associated. This represents a useful frequency of detection, considering that a number of the activities tested for represent extracellular functions, which may be less well expressed in a Gram-negative organism, and because the gene expression and protein export machinery of *B. cereus* and *E. coli* are likely to have numerous differences.

Genetic Analysis of BAC Clones: Identification of Loci Involved in Hemolysis, Esculin Hydrolysis, and Orange Pigment Production. Once a BAC clone expressing a particular activity is isolated, methods for rapid identification of the locus responsible for the activity will contribute to functional and physical analysis. We tested the possibility that transposon mutagenesis, combined with direct sequencing of BACs, could provide the appropriate information rapidly. Transposon mutants were isolated for a hemolysin-producing clone (BACB61; 40-kb insert), an esculin-hydrolyzing clone (BACB94; 80-kb insert), and an orange pigment producing clone (BACB142; 125-kb insert). These mutants were generated by infecting with lambda phage containing the Tn*pho*A marker or by electroporation with a suicide plasmid containing a mini-Tn*10*-Km transposon. These methods for plasmid mutagenesis were successful in all cases. The locus containing the transposon was identified by sequencing from the transposon into the flanking DNA by using primers specific for the ends of the transposon (Table 2).

The hemolytic activity of BACB61 likely is due to production of hemolysin II (26), because mutations that abolish

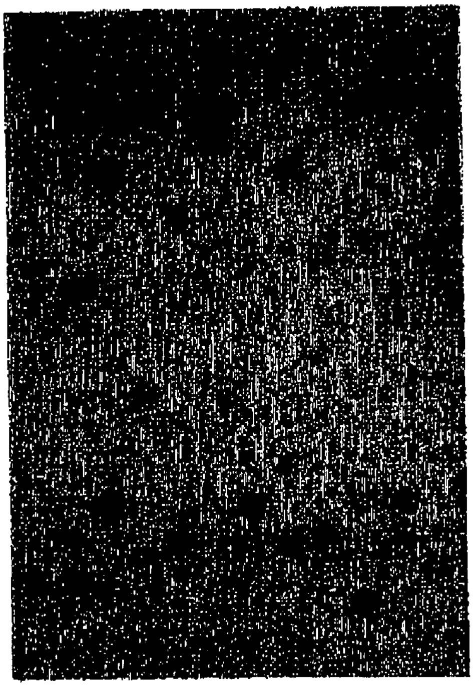


FIG. 2. Hybridization of BAC clones with a genomic *Not*I fragment. A 700-kb genomic *Not*I fragment from strain UW85 was used to probe a blot containing DNA from all BAC clones with inserts greater than 80 kb in size. These clones (total, 216) provide 4-fold coverage of the UW85 genome and can be screened on one blot.

Table 1. Expression of *B. cereus* activities in *E. coli*

Activity tested	No. of clones detected
Starch hydrolysis	0
Casein hydrolysis	0
Hemolysis	2
Esculin hydrolysis	2
Orange pigment	2
Ampicillin resistance	1
Zwittermicin A resistance	2
Lecithinase	2
Chitinase	0
Lipase	0

Table 2. Identification of *B. cereus* loci

Clone number	Activity	Transposon	Homology, %
BACB61	Hemolysis	Tn*phoA* Mini-Tn*10*-Km	YqiI (54% over 79 aa) HlyII (95% over 103 aa) Eag (52% over 51 aa)
BACB94	Esculin hydrolysis	Tn*phoA*	LevR (58% over 144 aa) CelR (47% over 127 aa)
BACB142	Orange pigment	Tn*phoA*	CatA (72% over 51 aa) CatV (40% over 62 aa)

Accession numbers for the above-mentioned loci: YqiI, Z99116; HlyII, U94743; Eag, X99724; LevR, M60105; CelR, U07818; CatA, M75944; CatV, M80796. References are cited in the text. Homology values indicate percent identity.

hemolytic activity are located in an ORF with homology to *hlyII* of another *B. cereus* strain (GenBank accession no. U94743). Before this analysis, strain UW85 was not known to contain this gene. Interestingly, mutations in the BAC clone that confer reduced but detectable levels of hemolytic activity also were isolated. Insertions producing this activity were in one of two ORFs: an ORF homologous to *B. subtilis* cell wall amidases (27) or an ORF encoding an S layer homology motif (28, 29). Restriction enzyme digestion analysis (not shown) indicated that the two loci were linked closely on BACB61, suggesting a possible functional or transcriptional linkage. Further analysis of this locus is required to understand its structure and contribution to hemolytic activity.

Two independently isolated mutations of BACB94, which confers the ability to hydrolyze esculin on *E. coli*, were found to be in an ORF with homology to the *B. subtilis* protein LevR. LevR is a transcription factor regulating expression of the levanase operon (30). It is likely that there is a β-glucosidase utilization operon present on BACB94 that is regulated by the LevR homolog, conferring esculin hydrolysis activity. Other genes similar to *levR* are postulated to regulate genes involved in cellobiose utilization [*celR* from *B. stearothermophilus* (31)] or β-glucoside utilization [*arbG* of *Erwinia chrysanthemi* (32)]. An alternative explanation is that the LevR homolog fortuitously activates the cryptic *bgl* operon in *E. coli*, because LevR has homology to BglG, the antiterminator protein that regulates *bgl* expression (33). These hypotheses now can be tested experimentally.

Two independently isolated transposon mutants were identified that disrupted pigment production in BACB142, the orange pigment-producing BAC clone. Both of these transposons disrupted a single ORF with homology to bacterial catalase enzymes (34, 35). These proteins contain a heme cofactor, which may be responsible for the orange color of the colonies carrying BACB142. Alternatively, the presence of genes on this BAC could disrupt or modify *E. coli* heme metabolism. Consistent with the latter idea, there appears to be an ORF directly upstream of the catalase ORF that has homology to bacterial ferrochelatases (data not shown), suggesting there may be alterations in heme metabolism in this strain. Overexpression of a heme biosynthetic enzyme from *B. stearothermophilus* in *E. coli* resulted in the colonies having a reddish color (36). Alternatively, there could be a gene downstream from the catalase gene on BACB142 that is responsible for the orange color and whose expression is altered by the presence of the transposon, although sequence analysis of the downstream region did not reveal any homology to known genes.

Genomic Comparisons by Using BAC Plasmids. Despite their large size and low copy number, sequencing directly from BACs is practical, as seen above. We generated sequence information at the ends of the insert DNA from the three characterized BACs, BACB61, BACB94, and BACB142, by using standard primers directed to vector sequences. This resulted in four of six cases in the identification of homologous sequences from *B. subtilis* (data not shown), reflecting the high information density of bacterial genomes. These data could be used to align a BAC to a sequenced genome from a related bacterium, if possible, or to select clones for complete sequencing. Construction, phenotypic analysis, and sequencing of insert ends with vector-directed primers of BAC libraries from several *Bacillus* species could provide a genomic overview of this group of organisms, which contain both industrially important and pathogenic species, and would provide a valuable resource for further genomic analysis

DISCUSSION

Our approach to functional genomics combines the utility of BAC libraries to access large, contiguous segments of DNA with the small genome size of prokaryotes and heterologous expression in *E. coli*. This merger results in a broadly applicable approach to the study of prokaryotic biology. Clearly, not every prokaryotic species is a candidate for whole genome sequencing, nor can we afford to develop genetic tools for each species of interest. BAC libraries offer a universal method to perform genetic, physical, and functional analyses of a prokaryotic genome without the need for an extensive investment in sequencing or specific methods development.

We predict that BAC libraries of bacterial genomes will yield new insights into prokaryotic biology, especially libraries of those species that, thus far, are poorly understood. Bacteria that are of significant biological interest for antibiotic production or ecological behavior, such as *B. cereus*, may be recalcitrant to the usual bacterial genetic and molecular techniques but can be studied via a BAC library approach.

BAC libraries of eukaryotic genomes typically contain thousands of clones (5). This number is required for sufficient coverage of large eukaryotic genomes. In contrast, prokaryotic genomes, such as that of *B. cereus*, require only a few hundred clones for equivalent coverage. For example, the BAC library of *Mycobacterium tuberculosis* required 68 BACs for a minimal overlap library of the 4.4-Mb genome, with one 150-kb gap (10). Yet, the same powerful techniques developed for eukaryotic BACs can be applied to prokaryotic BAC libraries, further increasing the utility of these libraries.

One of the advantages of BACs is that they appear to maintain heterologous DNA more stably than other cloning systems. This would be an advantage in cloning DNA from diverse microorganisms and might be especially relevant when gene expression from the clones is desired. The *M. tuberculosis* library has an average insert size of 70 kb, and inserts larger than 110 kb were not obtained (10), indicating that there might be species-specific limitations to some applications of BAC technology. We did not observe these problems with our *B. cereus* library, because inserts in our library were as large as 260 kb.

Given that six of ten of the activities screened for were detected in our library, we envision that diverse genes from bacteria can be cloned and analyzed successfully in BACs. We have not yet determined whether this frequency of detection of gene expression is an average level for BAC libraries of prokaryotic genomes. An answer to that will come from the construction and screening of more BAC libraries. Xu *et al.* (11) used pBeloBac11 to clone *Enterococcus faecalis* genes in *E. coli*, although in this library inserts were only 30–45 kb in size. Adequate levels of expression would be especially important when the activity of interest is hard to screen for or requires a large amount of genetic information for expression.

Though not broadly screened for in our library, BACs are the ideal tool for cloning and analysis of entire bacterial pathways, such as antibiotic biosynthesis pathways, biodegradative operons, or pathogenicity islands. Before our genetic analysis, it was unknown whether some activities such as hemolysis or orange pigment production were due to one or two genes or whether they represented biological activities of small molecules requiring a suite of genes for their production. Although we have applied BAC technology to *B. cereus*, a readily cultured organism, this approach will be even more powerful for analysis of the genomes of bacterial species in the environment, which may be accessed via BAC cloning, even if the cells themselves cannot be cultured at the present time (37, 38).

We thank Brett Tyler for introducing us to BACs and for a critical discussion that led to this work. We thank Amy Klimowicz and Daniel Cook for phenotypic screening of parts of the library, Elizabeth Stohl for screening the library for *zmaR*-containing clones, Brad Borlee for isolating the BAC894 Tn*phoA* mutants, the W. S. Reznikoff laboratory for the plasmid DNA isolation method, and Alan Bettermann for help with minipreps. Carlos Rios provided the λ-Tn*phoA* materials. pLOF-Km was provided by Sang-jin Suh. M.R.R. gratefully acknowledges support for a part of this research from a postdoctoral research fellowship (1 F32 GM18871-01A1) from the National Institutes of Health. This research was supported by the University of Wisconsin University–Industry Research program, The McKnight Foundation, Ariad Pharmaceuticals, and the National Institutes of Health (R03 AI41786-02.

1. Shizuya, H., Birren, B., Kim, U.-J., Mancino, V., Slepak, T., Tachiiri, Y. & Simon, M. (1992) *Proc. Natl. Acad. Sci. USA* 89, 8794–8797.
2. Monaco, A. P. & Larin, Z. (1994) *Trends Biotechnol.* 12, 280–286.
3. Kim, U.-J., Birren, B. B., Slepak, T., Mancino, V., Boysen, C., Kang, H.-L., Simon, M. I. & Shizuya, H. (1996) *Genomics* 34, 213–218.
4. Woo, S. S., Jiang, J., Gill, B. S., Paterson, A. H. & Wing, R. A. (1994) *Nucleic Acids Res.* 22, 4922–4931.
5. Venter, J. C., Smith, H. O. & Hood, L. (1996) *Nature (London)* 381, 364–366.
6. Mejia, J. E. & Monaco, A. P. (1997) *Genome Res.* 7, 179–186.
7. Yang, X. W., Model, P. & Heintz, N. (1997) *Nat. Biotechnol.* 15, 859–865.
8. Antoch, M. P., Song, E.-J., Chang, A.-M., Vitaterna, M. H., Zhao, Y., Wilsbacher, L. D., Sangoram, A. M., King, D. P., Pinto, L. H. & Takahashi, J. S. (1997) *Cell* 89, 655–667.
9. Kim, C. Y., Horrigan, S. K., Altenhofen, J. L., Arbieva, Z. H., Hoffman, R. & Westbrook, C. A. (1998) *Genome Res.* 8, 404–412.
10. Brosch, R., Gordon, S. V., Billault, A., Garnier, T., Eiglmeier, K., Soravito, C., Barrell, B. G. & Cole, S. T. (1998) *Infect. Immunol.* 66, 2221–2229.
11. Xu, Y., Jiang, L., Murray, B. E. & Weinstock, G. M. (1997) *Infect. Immunol.* 65, 4207–4215.
12. Handelsman, J., Raffel, S., Mester, E. H., Wunderlich, L. & Grau, C. R. (1990) *Appl. Environ. Microbiol.* 56, 713–718.
13. Harrell, L. J., Andersen, G. L. & Wilson, K. H. (1995) *J. Clin. Microbiol.* 33, 1847–1859.
14. Birren, B. & Lai, E. (1993) *Pulsed Field Gel Electrophoresis* (Academic, San Diego).
15. Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Lab. Press, Plainview, NY).
16. Gerhardt, P. (1994) *Methods for General and Molecular Bacteriology* (Am. Soc. Microbiol., Washington, DC).
17. Milner, J. L., Stohl, E. A. & Handelsman, J. (1996) *J. Bacteriol.* 178, 4266–4272.
18. Manoil, C. & Beckwith, J. (1985) *Proc. Natl. Acad. Sci. USA* 82, 8129–8133.
19. Kaufman, M. R. & Taylor, R. K. (1994) *Methods Enzymol.* 235, 426–448.
20. Herrero, M., de Lorenzo, V. & Timmis, K. (1990) *J. Bacteriol.* 172, 6557–6567.
21. Taylor, R. K., Manoil, C. & Mekalanos, J. J. (1989) *J. Bacteriol.* 171, 1870–1878.
22. Leclerc, G. J., Tartera, C. & Metcalf, E. S. (1998) *Infect. Immunol.* 66, 682–691.
23. Clarke, L. & Carbon, J. (1976) *Cell* 9, 91–99.
24. Zhu, H., Choi, S., Johnston, A. K., Wing, R. A. & Dean, R. A. (1997) *Fungal Genet. Biol.* 21, 337–347.
25. Boysen, C., Simon, M. I. & Hood, L. (1997) *Genome Res.* 7, 330–338.
26. Sinev, M. A., Budarina, Z. I., Gavrilenko, I. V., Tomashevskii, A. Y. & Kuz'min, N. V. (1993) *Mol. Biol.* 27, 753–760.
27. Kunst, F., Ogasawara, N., Moszer, I., Albertini, A. M., Alloni, G., Azevedo, V., Bertero, M. G., Bessieres, P., Bolotin, A., Borchert, S., *et al.* (1997) *Nature (London)* 390, 249–256.
28. Mesnage, S., Tosi-Couture, E., Mock, M., Gounon, P. & Fouet, A. (1997) *Mol. Microbiol.* 23, 1147–1155.
29. Lupas, A., Engelhardt, H., Peters, J., Santarius, U., Volker, S. & Baumeister, W. (1994) *J. Bacteriol.* 176, 1224–1233.
30. Débarbouillé, M., Martin-Verstraete, I., Klier, A. & Rapoport, G. (1991) *Proc. Natl. Acad. Sci. USA* 88, 2212–2216.
31. Lai, X. & Ingram, L. O. (1993) *J. Bacteriol.* 175, 6441–6450.
32. El Hassouni, M., Henrissat, B., Chippaux, M. & Barras, F. (1992) *J. Bacteriol.* 174, 765–777.
33. Amster-Choder, O. & Wright, A. (1993) *J. Cell. Biochem.* 51, 83–90.
34. Bol, D. K. & Yasbin, R. E. (1991) *Gene* 109, 31–37.
35. Haas, A., Brehm, K., Kreft, J. & Goebel, W. (1991) *J. Bacteriol.* 173, 5159–5167.
36. Chou, F.-L., Ohtsuka, M., Minowa, T., Yamasoto, K., Sakano, Y., Matsuzawa, H., Ohta, T. & Sakai, H. (1995) *Biosci. Biotech. Biochem.* 59, 1817–1824.
37. Stein, J. L., Marsh, T. L., Wu, K. Y., Shizuya, H. & DeLong, E. F. (1996) *J. Bacteriol.* 178, 591–599.
38. Handelsman, J., Rondon, M. R., Brady, S. F., Clardy, J. & Goodman, R. M. (1998) *Chem. Biol.* 5, R245–R249.

Vol. 186, No. 5

JOURNAL OF BACTERIOLOGY, Mar. 2004, p. 1571–1573
0021-9193/04/$08.00+0 DOI: 10.1128/JB.186.5.1571–1573.2004
Copyright © 2004, American Society for Microbiology. All Rights Reserved.

Characterization of Two Kinases Involved in Thiamine Pyrophosphate and Pyridoxal Phosphate Biosynthesis in *Bacillus subtilis*: 4-Amino-5-Hydroxymethyl-2-Methylpyrimidine Kinase and Pyridoxal Kinase

Joo-Hoon Park, Kristin Burns, Cynthia Kinsland, and Tadhg P. Begley*
Department of Chemistry and Chemical Biology, Cornell University, Ithaca, New York 14853

Received 29 May 2003/Accepted 10 October 2003

Two *Bacillus subtilis* genes encoding two proteins (currently annotated ThiD and YjbV) were overexpressed and characterized. YjbV has 4-amino-5-hydroxymethyl-2-methylpyrimidine and 4-amino-5-hydroxymethyl-2-methylpyrimidine pyrophosphate kinase activity and should be reannotated ThiD, and *B. subtilis* ThiD has pyridoxine, pyridoxal, and pyridoxamine kinase activity and should be reannotated PdxK.

The biosynthesis of thiamine pyrophosphate (TPP) involves the coupling of 4-amino-5-hydroxymethyl-2-methylpyrimidine pyrophosphate (HMP-PP) and 4-methyl-5-β-hydroxyethylthiazole phosphate (Thz-P) to form thiamine phosphate followed by a final phosphorylation (1). In addition to the de novo biosynthesis, microorganisms have developed several salvage pathways for the biosynthesis of TPP (Table 1). Thiamine from the growth medium is either phosphorylated by thiamine kinase or pyrophosphorylated by thiamine pyrophosphokinase (J. Melnick, E. Lis, J.-H. Park, H. Mori, C. Kinsland, J. Perkins, G. Schyns, A. Osterman, and T. P. Begley, submitted for publication). The pyrimidine and thiazole components can also be salvaged: thiazole is phosphorylated by thiazole kinase (2, 4, 6), HMP is phosphorylated to HMP-P by both ThiD and PdxK (3, 7, 10), and the phosphorylation of HMP-P is catalyzed by ThiD (5, 6, 7). Thus, ThiD has both a biosynthetic and a salvage function in thiamine biosynthesis. PdxK is able to phosphorylate a broad range of substrates, including HMP, pyridoxal (PL), pyridoxamine (PM), and pyridoxine (PN), and is a salvage enzyme in the biosynthesis of thiamine as well as that of PL phosphate (PLP).

A search of the *Bacillus subtilis* genomic database (http://genolist.pasteur.fr/SubtiList/index.html) shows homologues of *Escherichia coli* ThiD and PdxK named YjbV (1246149–1246961) and ThiD (3899983–3900795). They are both 271-amino-acid proteins. *yjbV* is located immediately downstream of the *thiOSGF* operon that is involved in Thz-P biosynthesis, while *thiD* is not clustered with any of the thiamine or PLP biosynthetic genes. *E. coli* PdxK shows 24 and 25% sequence identity with *B. subtilis* YjbV and ThiD, respectively, and *E. coli* ThiD shows 41 and 35% identity with *B. subtilis* YjbV and ThiD, respectively. The level of sequence homology between these two proteins is too high to allow the preferred substrate to be predicted for either protein. However, the occurrence of *yjbV* in the thiazole biosynthetic operon suggests that these proteins are incorrectly annotated and that YjbV might function as the *B. subtilis* HMP/HMP-P kinase. Here we report the overexpression of YjbV and ThiD from *B. subtilis* and the identification of the substrate preferences of the two proteins.

The amino acid sequences of *E. coli* ThiD and PdxK were obtained from the National Center for Biotechnology Information (http://www.ncbi.nlm.nih.gov/) and used with the SubtiList World Wide Web server for a BLAST search. For cloning *B. subtilis thiD* and *yjbV*, standard DNA restriction endonuclease digestion, ligation, and transformation methods were used (9). Genomic DNA and plasmid DNA were purified with a Wizard Plus SV genomic DNA kit and a DNA Miniprep kit, respectively (Promega). DNA fragments were separated by agarose gel electrophoresis, excised, and purified with a QIA

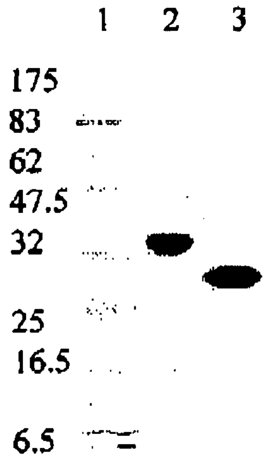


FIG. 1. SDS-PAGE (12%) analysis of purified *B. subtilis* ThiD and YjbV. Lane 1, molecular mass markers (in kilodaltons); lane 2, His-tagged ThiD; lane 3, His-tagged YjbV. Although ThiD and YjbV are predicted to have the same molecular mass, they migrate differently on the gel.

---

* Corresponding author. Mailing address: Department of Chemistry and Chemical Biology, Cornell University, Ithaca, NY 14853. Phone: (607) 255-7133. Fax: (607) 255-4137. E-mail: tpb2@cornell.edu.

TABLE 1. Microbial thiamine salvage enzymes

Enzyme	Microorganism	Substrate	Product	Reference(s)
ThiM	*Bacillus subtilis* *Escherichia coli* *Salmonella typhimurium*	Thz	Thz-P	2, 4, 6
PdxK	*Escherichia coli*	HMP	HMP-P	10
ThiD	*Escherichia coli*	HMP	HMP-P	7
ThiK	*Escherichia coli*	Thiamine	Thiamine phosphate	Melnick et al., submitted
ThiN	*Bacillus subtilis*	Thiamine	TPP	Melnick et al., submitted

quick gel extraction kit (Qiagen). pET-16b plasmid was obtained from Novagen. *E. coli* strain DH5α was used as a recipient for transformation during plasmid construction and for plasmid propagation and storage. *E. coli* BL21(DE3) was purchased from Novagen and used as a host strain for the overexpression of the proteins. A Perkin Elmer GeneAmp PCR System 2400 apparatus and Platinum Pfx DNA polymerase (Gibco Life Technologies) were used for PCR. *B. subtilis* CU1065 genomic DNA was used as a template for PCR. Primer synthesis and DNA sequencing were performed by the Bioresource Center at Cornell University. Primers introduced *Nde*I and *Xho*I restriction enzyme sites at the 5′ and 3′ ends, respectively.

For the overexpression and purification of ThiD and YjbV, their corresponding overexpression plasmids were transformed into competent *E. coli* BL21(DE3) cells and the transformed cells were grown at 37°C in Luria-Bertani medium containing 50 mg of ampicillin/liter. To induce the overexpression of proteins, isopropyl-β-D-thiogalactopyranoside (IPTG) was added to the culture (when the optical density at 595 nm reached 0.6) to achieve a final concentration of 1 mM. Culture growth was continued for 8 h at 28°C, after which the cells were harvested and stored at −80°C until further use. The proteins were purified according to a Qiagen protocol for the purification of His-tagged proteins. The eluted proteins were rapidly desalted using a PD-10 column (Amersham Pharmacia) because of instability under high-salt concentrations and stored in 5% glycerol at −80°C. ThiD was soluble and stable in 50 mM Tris buffer (pH 8), but YjbV solutions rapidly became turbid. The results of sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) with the purified proteins are shown in Fig. 1. Although the migration characteristics of the purified proteins were different, their molecular weights were confirmed by mass spectrometry (data not shown).

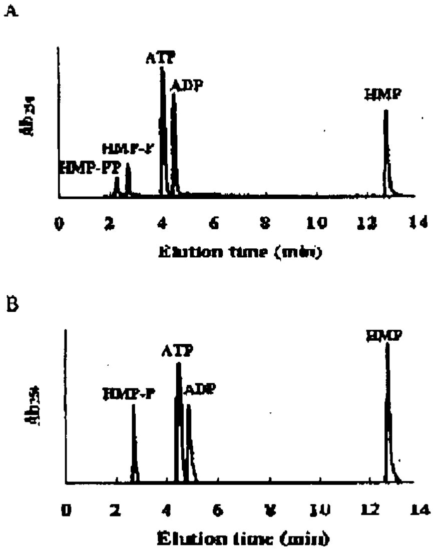


FIG. 2. HPLC analysis of the ThiD and YjbV catalyzed reactions. (A) YjbV catalyzed phosphorylation of HMP and HMP-P. (B) ThiD catalyzed phosphorylation of HMP.

The reaction mixtures for *B. subtilis* ThiD and YjbV enzymatic assays contained 1 mM ATP, 1 mM HMP, 2 mM $MgCl_2$, and 40 μg of enzyme in 100 μl of 50 mM Tris-HCl (pH 8). After incubation at 37°C for 10 min, the reaction was quenched by the addition of 100 μl of 10% trichloroacetic acid and centrifuged to remove proteins. A total of 20 μl of the reaction mixture was analyzed by high-pressure liquid chromatography (HPLC) (Supelcosil LC-18-T) (15- by 4.6-mm column). The elution conditions were as follows: flow rate, 1 ml/min; elution time, 0 to 20 min; elution buffer, 100% of 0.1 M potassium phosphate (pH 6.6). To conduct a competition assay, ThiD was incubated with all four substrates (0.3 mM concentrations each of HMP, PL, PM, and PN) for 30 min under the conditions described above (except that 2 mM ATP was used and the reaction mixture was analyzed by HPLC).

For kinetic studies, ADP produced by the kinase activity of ThiD or YjbV was assayed using a pyruvate kinase lactate dehydrogenase-coupled system (which uses ADP and NADH as substrates). The consumption of NADH by this coupled system can be measured by monitoring the decrease in absorbance at 340 nm (7). Pyruvate kinase, lactate dehydrogenase, phosphoenolpyruvate, NADH, and PL were purchased from Sigma. HMP was synthesized as previously described (8). The assay mixture for the kinetic analysis of ThiD in the presence of HMP or PL contained saturating concentrations of ATP (5 mM), 30 to 400 μM HMP (or 30 to 300 μM PL), 10 mM $MgCl_2$, 50 mM KCl, 0.2 mM NADH, 1 mM phosphoenolpyruvate, 8 units of pyruvate kinase/ml, and 10 units of lactate dehydrogenase/ml in 0.6 ml of 50 mM Tris-HCl (pH 8). Addition of ThiD to achieve a final concentration of 6.7 μM initiated the reactions, which were then monitored over 5 min for NADH consumption at 340 nm.

HPLC analysis of the reaction mixture containing *B. subtilis*

TABLE 2. Kinetic parameters for substrate phosphorylation by *B. subtilis* ThiD

Substrate	K_m (μM)	k_{cat} (s-1)	k_{cat}/K_m (s-1 μM-1)
HMP	2,030	0.36	1.8×10^{-4}
PL	46.6	0.032	6.9×10^{-4}

YjbV showed the appearance of two new peaks corresponding to HMP-P and HMP-PP (Fig. 2A). The reaction mixture containing *B. subtilis* ThiD showed only one pyrimidine product peak, which corresponds to HMP-P (Fig. 2B). In addition to the phosphorylation of HMP, *B. subtilis* ThiD was able to phosphorylate PL, PM, and PN, producing PLP, PMP, and PNP, respectively. Under similar conditions, YjbV did not catalyze the phosphorylation of these compounds (data not shown). A competition assay using the substrates of ThiD revealed a preference for PL followed by HMP, PN, and PM (8:2.4:1.1:1 product ratios). The kinetic parameters for *B. subtilis* ThiD are shown in Table 2. The kinetic parameters of *B. subtilis* YjbV could not be determined, because the reaction mixture became turbid immediately after the reaction began.

Overall our results indicate that *B. subtilis* YjbV has HMP/HMP-P kinase activity and should be reannotated ThiD (i.e., the name should be changed from YjbV to ThiD) and that *B. subtilis* ThiD has PN/PL/PM/HMP kinase activity and should be reannotated PdxK (i.e., from ThiD to PdxK).

This research was supported by a grant from NIH (DDK44083).

REFERENCES

1. **Begley, T. P., D. M. Downs, S. E. Ealick, F. W. McLafferty, A. P. G. M. Van Loon, S. Taylor, N. Campobasso, H. J. Chiu, C. Kinsland, J. J. Reddick, and J. Xi.** 1999. Thiamin biosynthesis in prokaryotes. Arch. Microbiol. 171:293–300.
2. **Campobasso, N., I. I. Mathews, T. P. Begley, and S. E. Ealick.** 2000. Crystal structure of 4-methyl-5-β-hydroxyethylthiazole kinase from *Bacillus subtilis* at 1.5 Å resolution. Biochemistry 39:7868–7877.
3. **Mizote, T., and H. Nakayama.** 1989. Purification and properties of hydroxymethylpyrimidine kinase from *Escherichia coli*. Biochim. Biophys. Acta 991:109–113.
4. **Mizote, T., and H. Nakayama.** 1989. The *thiM* locus and its relation to phosphorylation of hydroxyethylthiazole in *Escherichia coli*. J. Bacteriol. 171:3228–3232.
5. **Mizote, T., M. Tsuda, D. D. S. Smith, H. Nakayama, and T. Nakazawa.** 1999. Cloning and characterization of the *thiD/J* gene of *Escherichia coli* encoding a thiamin-synthesizing bifunctional enzyme, hydroxymethylpyrimidine kinase/phosphomethylpyrimidine kinase. Microbiology 145:495–501.
6. **Petersen, L. A., and D. M. Downs.** 1997. Identification and characterization of an operon in *Salmonella typhimurium* involved in thiamine biosynthesis. J. Bacteriol. 179:4894–4900.
7. **Reddick, J. J., C. Kinsland, R. Nicewonger, T. Christian, D. M. Downs, M. E. Winkler, and T. P. Begley.** 1998. Overexpression, purification and characterization of two pyrimidine kinases involved in the biosynthesis of thiamin: 4-amino-5-hydroxymethyl-2-methylpyrimidine kinase and 4-amino-5-hydroxymethyl-2-methylpyrimidine phosphate kinase. Tetrahedron 54:15983–15991.
8. **Reddick, J. J., R. Nicewonger, and T. P. Begley.** 2001. Mechanistic studies on thiamin phosphate synthase: evidence for a dissociative mechanism. Biochemistry 40:10095–10102.
9. **Sambrook, J., and D. W. Russell.** 2001. Molecular cloning: a laboratory manual, 3rd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
10. **Yang, Y., G. Zhao, and M. E. Winkler.** 1996. Identification of the pdxK gene that encodes pyridoxine (vitamin B6) kinase in *Escherichia coli* K-12. FEMS Microbiol. Lett. 141:89–95.

CbiZ, an amidohydrolase enzyme required for salvaging the coenzyme B_{12} precursor cobinamide in archaea

Jesse D. Woodson and Jorge C. Escalante-Semerena*

Department of Bacteriology, University of Wisconsin, Madison, WI 53726-4087

Edited by Ralph S. Wolfe, University of Illinois at Urbana–Champaign, Urbana, IL, and approved December 31, 2003 (received for review September 15, 2003)

The existence of a pathway for salvaging the coenzyme B_{12} precursor dicyanocobinamide (Cbi) from the environment was established by genetic and biochemical means. The pathway requires the function of a previously unidentified amidohydrolase enzyme that converts adenosylcobinamide to adenosylcobyric acid, a bona fide intermediate of the *de novo* coenzyme B_{12} biosynthetic route. The *cbiZ* gene of the methanogenic archaeon *Methanosarcina mazei* strain Gö1 was cloned, was overproduced in *Escherichia coli*, and the recombinant protein was isolated to homogeneity. HPLC, UV-visible spectroscopy, MS, and bioassay data established adenosylcobyric as the corrinoid product of the CbiZ-catalyzed reaction. Inactivation of the *cbiZ* gene in the extremely halophilic archaeon *Halobacterium* sp. strain NRC-1 blocked the ability of this archaeon to salvage Cbi. *cbiZ* function restored Cbi salvaging in a strain of the bacterium *Salmonella enterica*, whose Cbi-salvaging pathway was blocked. The salvaging of Cbi through the CbiZ enzyme appears to be an archaeal strategy because all of the genomes of B_{12}-producing archaea have a *cbiZ* ortholog. Reasons for the evolution of two distinct pathways for Cbi salvaging in prokaryotes are discussed.

Among vitamins and coenzymes, cobamides (e.g., coenzyme B_{12}) are unique for their structural complexity. Not surprisingly, *de novo* synthesis of cobamides requires a great deal of genetic information, which is only found in prokaryotes (1–3). The majority of the work on B_{12} biosynthesis has been performed in bacteria (4–7). In addition to a *de novo* pathway, bacteria also possess a conserved salvaging pathway for the precursor dicyanocobinamide (Cbi), which is a stable precursor, but is not a true intermediate of the *de novo* pathway (Fig. 1 and refs. 8 and 9). In bacteria, Cbi salvaging requires attachment of the upper ligand 5′-deoxyadenosine to yield adenosylcobinamide (AdoCbi) (10–13), followed by phosphorylation of AdoCbi to yield AdoCbi-phosphate (AdoCbi-P), which is a true intermediate of the *de novo* biosynthetic pathway (8). The latter reaction is catalyzed by the kinase activity of a bifunctional ATP:AdoCbi kinase, GTP:AdoCbi-GDP guanylyltransferase enzyme (CobU) in *Salmonella enterica*), which is conserved in cobamide producing bacteria (14–18).

It is clear that some archaea require and synthesize cobamides to live. However, our understanding of how archaea salvage Cbi is limited (8, 19). Analysis of the available archaeal genome sequences revealed the absence of an ortholog of the bacterial bifunctional CobU enzyme, and we recently reported the identification of the gene encoding the nonorthologous replacement of only the GTP:AdoCbi-GDP guanylyltransferase activity in archaea (8, 19). To the best of our knowledge, there have been no reports of ATP:AdoCbi kinase activity in any archaeon. More recent work from our laboratory showed that *Halobacterium* sp. strain NRC-1 can salvage Cbi, and that Cbi salvaging requires the activity of the cobyric acid synthase (CbiB) enzyme that catalyzes the last step of the *de novo* corrin ring biosynthetic pathway (20). These findings were consistent with the existence of an alternative Cbi-salvaging pathway in which AdoCbi is converted to adenosylcobyric acid (AdoCby) by a previously unidentified amidohydrolase enzyme. Here, we report the identification of the *cbiZ* gene as the one encoding the AdoCbi amidohydrolase in the methanogenic archaeon *Methanosarcina mazei* strain Gö1, and in the extremely halophilic archaeon *Halobacterium* sp. strain NRC-1.

Materials and Methods

Strains and Plasmids. Descriptions of the genotypes of strains and plasmids used in this work, as well as detailed descriptions of plasmid constructions, can be found in *Supporting Methods* and Table 1, which are published as supporting information on the PNAS web site. A diagram of the *Halobacterium* sp. NRC-1 DNA included in the most relevant plasmids are included in Fig. 8, which is published as supporting information on the PNAS web site.

Chemicals, Growth Media, Growth Conditions, and Assessment of Viability. Except where noted, all chemicals used in this work were high-purity, commercially available compounds. When added to the medium, corrinoids were present at 100 pM for *Halobacterium* sp. strain NRC-1 studies and 15 nM for *S. enterica* studies. All corrinoids were added in their cyano form. AdoCbi was synthesized as described (8); $(CN)_2Cbi$ and CNB_{12} [also known as CNCbl (cobalamin)] were purchased from Sigma. $(CN)_2Cbi$-GDP was synthesized as described (15); $(CN)_2Cby$ was a gift from P. Renz (Universität-Hohenheim, Stuttgart, Germany); 5-fluoroorotic acid was purchased from Zymo Research (Orange, CA); and mevinolin was purchased from LKT Laboratories (St. Paul) (*R*)-1-amino-2-propanol (AP) was purchased from Sigma.

Bacterial Strains Used for Protein Overproduction. Overproduction of native *M. mazei* CbiZ protein was performed in *Escherichia coli* strain BL21(λ DE3)-RIL (Stratagene). Overproduction of the CbiZ-chitin binding protein fusion protein was performed in *E. coli* strain ER2566 (New England Biolabs).

Growth Studies. Cultures of strains of *Halobacterium* sp. strain NRC-1 and *S. enterica* were grown as described (20). The only modification was the addition of 5,6-dimethylbenzimidazole (3

This paper was submitted directly (Track II) to the PNAS office.

Abbreviations: Cbi, dicyanocobinamide; CbiB, cobyric acid synthase; AdoCbi, adenosylcobinamide; Cby, cobyric acid; AdoCby, adenosylCby; AdoCbi-P, adenosylCbi-phosphate; AP, (R)-1-amino-2-propanol; CobU, GTP:AdoCbi-[illegible] guanylyltransferase enzyme; CobD, L-threonine decarboxylase.

Data deposition: The sequences reported in this paper have been deposited in the GenBank database (accession nos. GI15790554 [*Halobacterium* sp. strain NRC-1 ORF Vng1583C] and GI2122[illegible] [*Methanosarcina mazei* ORF Mm0173]).

*To whom correspondence should be addressed at: Department of Bacteriology, University of Wisconsin, 264 Enzyme Institute, 1710 University Avenue, Madison, WI 53726-4087. E-mail: escalante@bact.wisc.edu.

© 2004 by The National Academy of Sciences of the USA

www.pnas.org/cgi/doi/10.1073/pnas.0305939101

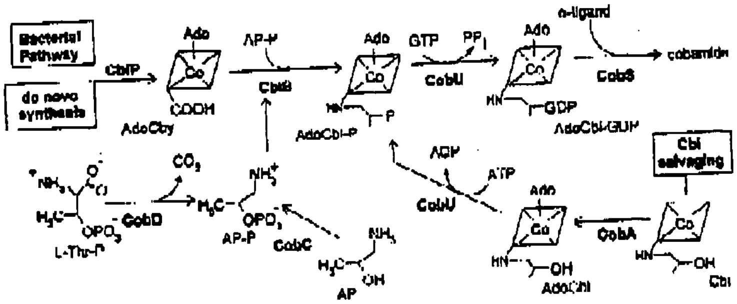


Fig. 1. Late steps of cobamide biosynthesis in bacteria. Intermediates are indicated below structures. AP-P, AP phosphate; CobS, cobalamin (5′-P) synthase; CobC, L-threonine kinase; CobA, ATP:co(I)rrinoid adenosyltransferase.

μM) and AP (10 mM) to the *S. enterica* medium. All plasmids introduced into *S. enterica* were first passed through a restriction-deficient strain (21).

Generation of Cell-Free Extracts Enriched with CbiZ Protein. The wild-type allele of *M. mazei cbiZ* was overproduced from the plasmid pMmCBIZ1 in the overproducing strain of *E. coli* BL21(λ DE3)-RIL (Stratagene). An overnight, 5-ml culture of the overproduction strain carrying pMmCBIZ1 was grown in LB broth containing ampicillin (100 μg/ml) and chloramphenicol (20 μg/ml). The overnight culture was used to inoculate 400 ml of fresh medium. The culture was grown at 30°C with shaking to a cell density of OD_{650} = 0.55, at which point isopropyl-β-D-thiogalactopyranoside was added to a final concentration of 500 μM. After the addition of isopropyl-β-D-thiogalactopyranoside, the culture was incubated under the same conditions for an additional 3 h. Cells were harvested by centrifugation for 20 min at 10,000 × *g* at 4°C by using a Beckman Coulter J21 centrifuge. Cells were resuspended in 4 ml of 50 mM Hepes buffer, pH 7.5, containing NaCl (100 mM) and DTT (5 mM). Cells were broken by two passes through a French press at ≈1 × 10^5 kPa by using a chilled pressure cell. Cell lysate was clarified by centrifugation for 30 min at 18,000 × *g* at 4°C using a Beckman-Coulter Avanti J-251 centrifuge. Soluble extract was dialyzed against 1 liter of the resuspension buffer (1:250) in a Slidalyzer (Pierce) cassette (molecular weight cutoff of 10,000) with two buffer changes. As negative control, the same procedure was used to generate cell-free extract from cells harboring the control plasmid pT7–7.

***In Vitro* CbiZ Amidohydrolase Activity Assay.** AdoCbi amidohydrolase activity assays were performed in 100-μl volumes containing 65 μg of cell-free extract protein, 50 mM Na-Hepes buffer, pH 7.5, containing DTT (5 mM) and AdoCbi (30 μM). Reactions were incubated at 37°C for 2 h in dim light and were heat-inactivated at 65°C for 20 min. As a negative control, heat inactivated cell free extract was prepared by incubating the extract at 65°C for 20 min. When the assay was performed with highly purified CbiZ, the reactions contained 150 μM AdoCbi, and 4 μg of CbiZ protein was used in lieu of crude cell-free extract.

Detection of Cby. The presence of Cby in reaction mixtures was assessed by means of a bioassay. For this purpose, *S. enterica* strain JE824 (*metE205 cobU330*) carrying plasmid pCOBY10 ($cobY^+$) was used as indicator strain. Five microliters of 1:10 dilutions of a reaction mixture was spotted onto an agar overlay containing cells of strain JE824 carrying plasmid pCOBY10. As controls, 5 pmol of authentic $(CN)_2$Cbi and $(CN)_2$Cby were also spotted onto the overlay. Minimal no-carbon E medium was supplemented with glucose, $MgSO_4$, 1,2-propanediol (to induce transcription of the *cbi* operon) (22), and ampicillin (25 μg/ml). Medium was incubated aerobically at 37°C for 24 h. Under aerobic conditions, *de novo* corrin ring biosynthesis is blocked in *S. enterica*, hence making growth dependent on corrinoid intermediates. Cell growth around the application sites indicated the presence of Cby in reaction mixtures.

Overproduction and Purification of Recombinant CbiZ Protein. *M. mazei* CbiZ protein fused to a C-terminal chitin-binding protein tag was overproduced by using plasmid pMmCBIZ5 in the overproducing strain of *E. coli* ER2566 (Stratagene). One milliliter of an overnight culture of the overproducing strain carrying plasmid pMmCBIZ5 in LB broth containing ampicillin (100 μg/ml) was used to inoculate two 500 ml batches of fresh medium. Cultures were grown at 30°C with shaking to a cell density of OD_{600} = 0.55, at which point isopropyl-β-D-thiogalactopyranoside was added to a final concentration of 300 μM. The isopropyl-β-D-thiogalactopyranoside containing culture was incubated under the same conditions for an additional 3 h. Cells were harvested by centrifugation for 20 min at 8,000 × *g* at 4°C. Cells were resuspended in 20 ml of 20 mM Hepes buffer, pH 7.5, containing NaCl (500 mM), EDTA (0.1 mM), and Triton X-100 (0.1% vol/vol). Cells were broken by two passes through a French press at ≈1.6 × 10^5 kPa by using a chilled pressure cell. Cell lysate was clarified by centrifugation for 30 min at 18,000 × *g* at 4°C. CbiZ protein was purified on chitin beads (New England Biolabs) according to the manufacturer's instructions. The chitin tag was removed from the protein by soaking the chitin beads in the same buffer containing 30 mM DTT for 20 h at 4°C. After purification, 5 ml of the enzyme was dialyzed by using snakeskin-pleated dialysis tubing (molecular weight cutoff of 10,000) (Pierce) at 4°C against 1 liter (1:200) of 50 mM Hepes buffer, pH 7.5, containing NaCl (100 mM) and DTT (5 mM). After two buffer changes, purity of the protein was assessed by SDS/PAGE (23) after staining with Coomassie brilliant Blue R-250 (24). CbiZ protein was stored at −80°C in this buffer after flash-freezing with liquid N_2.

Corrinoid Analysis. Corrinoids present in the reaction mixture were derivatized to their cyano form by adding 10 μl of 100 mM KCN and incubating in the light at room temperature for 30 min. Samples were filtered by using Corning Spin-X centrifuge filters. Corrinoids were separated by using a Waters HPLC system equipped with a Luna (Phenomenex) 5-μ C18 column (150 × 4.6 mm) developed with a modification of the system reported elsewhere (25) at a flow rate of 1 ml/min^{-1}. The column was equilibrated with a buffer system containing 98% A/2% B. One minute after injection of the sample, the column was developed for 5 min with a linear gradient until the final composition of the buffer system was 75% A/25% B. A second linear gradient of 15 min developed the column to a final buffer composition of 65% A/35% B; solvent A = 100 mM phosphate buffer, pH 6.5, 10 mM KCN; solvent B = 100 mM phosphate buffer, pH 8.0, 10 mM KCN:acetonitrile (1:1). Corrinoids were identified by their spectra by using a Waters photodiode array detector. Authentic $(CN)_2$Cbi and $(CN)_2$Cby were used as standards.

MS. The HPLC-purified product of the CbiZ reaction was dried under vacuum by using a Savant concentrator, was resuspended in 5 ml of ddH_2O, and was loaded onto a 1-ml LiChroprep RP-8 (EM Separations) column equilibrated with ddH_2O. The column was washed with 30 ml of ddH_2O, and corrinoid was eluted with 5 ml of methanol. The eluted sample was dried under vacuum, resuspended in 1 ml of ddH_2O, filtered in Spin-X filters, and again dried under vacuum. The sample was submitted for analysis to the MS facility at the University of Wisconsin-Madison Biotechnology Center. The mass spectrum was obtained by using a Bruker Daltonics (Billerica, MA) BIFLEX III matrix-assisted laser desorption/ionization-time-of-flight mass spectrometer.

Results

Identification of the Gene Encoding the Archaeal AdoCbi Amidohydrolase Enzyme. A comparative genomics approach was used to identify genes that could encode AdoCbi amidohydrolase activity (i.e., putative hydrolases or conserved hypothetical protein ORFs near known Cbi genes). This approach allowed for the identification of an uncharacterized conserved protein sometimes flanking identified ORFs involved in the late steps of cobamide biosynthesis. ORF Vng1583C (gi 15790554) (hereafter referred to as *cbiZ*) of *Halobacterium* sp. strain NRC-1 is the last gene of a putative four gene operon (*cobSYDcbiZ*) encoding orthologs of bacterial functions known to catalyze late steps of coenzyme B_{12} synthesis (Fig. 8).

cbiZ (ORF Vng1583C) Function Is Required for Cbi Salvaging in *Halobacterium*. To determine whether CbiZ was required for Cbi salvaging, a derivative of strain JE6738 (Δ*cbiP*) was constructed. Strain JE6812 (Δ*cbiP* Δ*cbiZ*) carried an in-frame deletion of the gene encoding the Cbi acid synthase (CbiP), the enzyme that catalyzes the second-to-last step of the corrin ring biosynthesis. A mutation in *cbiP* blocks *de novo* corrin ring synthesis in *Halobacterium* (20), thus demanding salvaging of precursors present in the medium. Strain JE6812 (Δ*cbiP* Δ*cbiZ*) was tested for its ability to salvage different corrinoids. Like strain JE6738 (Δ*cbiP cbiZ*$^+$), strain JE6812 (Δ*cbiP* Δ*cbiZ*) failed to grow in chemically defined medium lacking corrinoids (Fig. 2*A*, ●), and the addition of $(CN)_2$Cby [a derivatized *de novo* pathway intermediate (Fig. 1)] restored wild-type growth (Fig. 2*A*, ▲; doubling time = 27 h). Addition of either $(CN)_2$Cbi-GDP [a derivatized *de novo* pathway intermediate (Fig. 1)] or CNB_{12} also restored wild-type growth of both strains (data not shown). However, whereas the addition of $(CN)_2$Cbi to the medium allowed wild-type growth of JE6738 (Δ*cbiP*) (doubling time = 24 h), it did not support growth of strain JE6812 (Δ*cbiP* Δ*cbiZ*) (Fig. 2*A*, ○ vs. ◇). These data established a strong correlation between the loss of *cbiZ* function and a block in Cbi salvaging under conditions that demanded salvaging of this precursor.

***cbiZ* Function Is Necessary and Sufficient for Cbi Salvaging in *Halobacterium*.** The observed block in Cbi salvaging in strain JE6812 (Δ*cbiP* Δ*cbiZ*) was corrected when a wild-type allele of *cbiZ* was reintroduced into the chromosome. Strain JE7210 (Δ*cbiP* Δ*cbiZ ura3::cbiZ*$^+$) grew in chemically defined medium supplemented with Cbi with a doubling time of 22 h (Fig. 2*B*, □), but did not grow without corrinoids due to the lack of *cbiP* function (Fig. 2*B*, ●). Hence, *cbiZ*$^+$ was necessary and sufficient to restore Cbi salvaging in the mutant strain. To demonstrate that *cbiZ* function was not required for *de novo* corrin ring synthesis, the *de novo* pathway of strain JE6812 (Δ*cbiP* Δ*cbiZ*) was restored by reintroducing a wild-type *cbiP* allele into the chromosome. Strain JE7002 (Δ*cbiP* Δ*cbiZ ura3::cbiP*$^+$) grew in medium without any corrinoids (Fig. 2*B*, ▼; doubling time = 27 h). It was concluded that *cbiP* function was necessary and sufficient to restore *de novo* cobamide synthesis in strain JE7002 even in the absence of *cbiZ*.

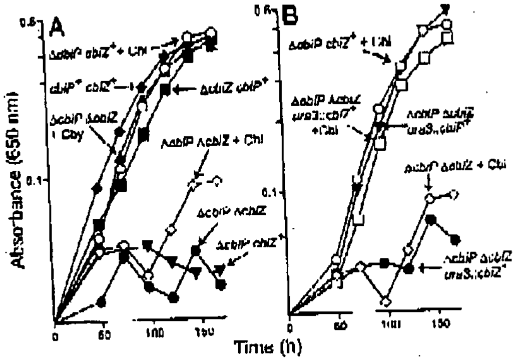


Fig. 2. Nutritional analyses of *Halobacterium* strains. Shown is B_{12}-dependent growth of *Halobacterium* strains in defined liquid medium at 37°C. Strains are indicated by their genotype. Corrinoids added to the medium are indicated next to each genotype. Strains used were JE6738, Δ*cbiP cbiZ*$^+$; JE6811, *cbiP*$^+$ Δ*cbiZ*; JE6812, Δ*cbiP* Δ*cbiZ*; JE7002, Δ*cbiP* Δ*cbiZ ura3::cbiP*$^+$; and JE7210, Δ*cbiP* Δ*cbiZ ura3::cbiZ*$^+$. All corrinoids were present in the medium at 100 pM.

***cbiZ* Is Not Involved in *de Novo* Biosynthesis in *Halobacterium*.** To determine whether strain JE6811 (Δ*cbiZ*) was deficient in *de novo* cobamide biosynthesis, growth was assessed in chemically defined liquid medium under conditions where cobamides were essential for growth. The kinetics of growth of strain JE6811 (Δ*cbiZ*) in chemically defined medium lacking corrinoids was very similar to that of the wild-type strain (Fig. 2*A*, ■ vs. ◆ with doubling times of 30 and 34 h, respectively.

An Archaeal *cbiZ* Gene Restores Cbi Salvaging in an *S. enterica* Mutant Strain. To lend support to the conclusion that CbiZ was involved in Cbi salvaging in archaea, the ability of an archaeal *cbiZ*$^+$ gene to complement an *S. enterica* strain defective in Cbi salvaging was tested. In previous experiments in our laboratory, *Halobacterium* sp. strain NRC-1 genes have failed to complement *S. enterica* mutants, presumably because of the severe difference in internal salt concentrations. Instead, we focused on the archaeal ortholog of *cbiZ* [ORF Mm0173 (gi 21226275)] in *M. mazei* Göl, a mesophilic, methanogenic archaeon whose genes have been successfully expressed in *S. enterica* (20). The CbiZ-dependent Cbi salvaging pathway is also expected to exist in *M. mazei* Göl, because it is predicted to have orthologs to proteins required for this pathway (20, 26).

S. enterica strain JE824 (*metE205 cobU330*) was used to test for *MmcbiZ*$^+$ function. Growth of this strain depends on cobamide-dependent methylation of homocysteine by the cobamide-dependent methionine synthase enzyme (27). Because the *de novo* pathway in *S. enterica* is inactive under aerobic conditions, growth in defined medium lacking methionine requires our

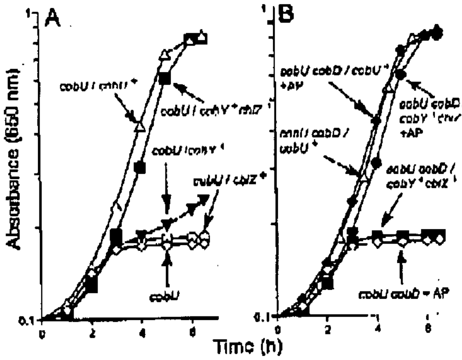


Fig. 3. Nutritional analyses of *S. enterica* strains. Cbi salvaging-dependent growth of *S. enterica* strains in chemically defined liquid medium at 37°C. (*A*) All strains were derivatives of strain JE824 (*metE205 cobU330*). (*B*) All strains were derivatives of strain JE6894 (*metE205 cobU330 cobD1272*). Strains are indicated by their genotype. Plasmids used were pT7-7, control vector; pJO52, *cobU*$^+$; pCOBY10, *cobY*$^+$; pMmCBIZ1, *cbiZ*$^+$; and pMmCBIZ2, *cbiZ*$^+$ *cobY*$^+$. In all cases, (CN)$_2$Cbi was added to 15 nM.

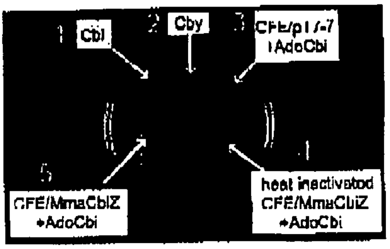


Fig. 4. Bioassay for the detection of Cby synthesized *in vitro* by cell-free extract of *E. coli* cells overproducing CbiZ protein. Shown is the response of indicator strain JE824 (*metE205 cobU330*) with plasmid pCOBY10 (*cobY*$^+$) to 5 μl of 1:10 dilutions of the deproteinated reaction mixtures and 5 pmol of standards. Growth around the area of application indicates the presence of Cby in the reaction mixture. Shown are the (CN)$_2$Cbi standard (spot 1), the (CN)$_2$Cby standard (spot 2), vector-only control (spot 3), heat-inactivated control (spot 4), and the complete reaction mixture (spot 5).

rinoid salvaging. Mutation *cobU330* eliminates both the ATP:AdoCbi kinase and the GTP:AdoCbi-P guanylyltransferase activities and blocks *de novo* corrin ring biosynthesis and Cbi salvaging in this bacterium (28). To restore CbiZ, Cbi-dependent growth of strain JE824, an NTP:AdoCbi-P nucleotidyltransferase was provided. For this purpose, the *M. mazei cobY*$^+$ gene was introduced into strain JE824. Plasmid pJO52 (*cobU*$^+$) was used as positive control, whereas plasmid pT7–7 was used as vector-only, negative control. Plasmids pMmCBIZ1 (*cbiZ*$^+$), pMmCBIZ2, (*cbiZ*$^+$ *cobY*$^+$), or pCOBY10 (*cobY*$^+$) were introduced into strain JE824. Resulting strains were grown aerobically in medium supplemented with Cbi. Under the conditions used, growth depended on Cbi salvaging. Cbi-dependent growth was only observed when either *S. enterica cobU*$^+$ (Fig. 3*A*, △) or *M. mazei cbiZ*$^+$ and *cobY*$^+$ were provided *in trans* (Fig. 3*A*, ■). These data supported the conclusion that *cbiZ* function was required for Cbi salvaging.

***cbiZ* Restores Cbi Salvaging Via a Pathway Different from the One Found in Bacteria.** Although CbiZ restored Cbi salvaging in strain JE824, these results did not shed any insights into how Cbi was salvaged. To identify the entry point for Cbi, we used an *S. enterica* strain carrying a mutation in the L-threonine decarboxylase (CobD) enzyme. A block at this step in the pathway would not affect Cbi salvaging if the entry point was AdoCbi-P, as expected if CobU were functional (Fig. 1). If, however, Cbi was converted to an earlier intermediate, CobD function would be required for Cbi salvaging.

Plasmid pMmCBIZ2 (*cobY*$^+$ *cbiZ*$^+$) failed to restore Cbi salvaging in strain JE6984 (*metE205 cobU330 cobD1272*) (Fig. 3*B*, ■); the control strain carrying plasmid pJO52 (*cobU*$^+$) salvaged Cbi (Fig. 3*B*, △). Cbi salvaging by the strain carrying plasmid pMmCBIZ2 was restored when AP was added to the medium (Fig. 3*B*, ●). This result was expected because the addition of AP compensates for the lack of CobD function (9, 29). These results were consistent with the idea that CbiZ-dependent Cbi salvaging occurs via an alternative pathway that converts Cbi to an intermediate before AdoCbi-P.

Cell-Free Extracts Enriched for CbiZ Contain an AdoCbi Amidohydrolase Activity. We considered the possibility that CbiZ was an amidohydrolase enzyme that converted Cbi to Cby by removing the AP propanol moiety of Cbi. To test this idea, *M. mazei* CbiZ was overproduced from plasmid pMmCBIZ1 (*cbiZ*$^+$) in *E. coli* overproducing strain BL21(λDE3)-RIL. AdoCbi was incubated with cell-free extract enriched for CbiZ, and amidohydrolase activity was measured by using a bioassay that detected the presence of Cby in the reaction mixture. If Cbi were converted to Cby by CbiZ, the cobamide auxotrophy of strain JE824 carrying plasmid pMmCOBY10 (*cobY*$^+$) would be corrected, resulting in growth around the application point of the sample. Cby synthesis was only detected in reaction mixtures that contained cell-free extract enriched with CbiZ protein (Fig. 4, spot 5), suggesting the CbiZ protein had AdoCbi amidohydrolase enzyme activity.

CbiZ-Dependent Conversion of Cbi to Cby. The CbiZ protein was purified to homogeneity by using a C-terminal chitin-binding protein tag, which was subsequently cleaved (data not shown). Purified CbiZ enzyme (>95 homogeneity by SDS/PAGE) was tested for AdoCbi amidohydrolase activity by incubating the protein with AdoCbi and monitoring the formation of the product Cby by using HPLC protocols described above. A signal for Cby was clearly detectable in the complete reaction mixture (Fig. 5*A*), but was absent when CbiZ protein was inactivated before incubation with the substrate (Fig. 5*B*). Under these conditions, a specific activity of 6.4 μ mol per min per mg of protein was calculated. When (CN)$_2$Cbi was used as substrate, the specific activity was reduced 3 fold (2.1 μmol per min per mg of protein). To demonstrate the reduced activity was not due to

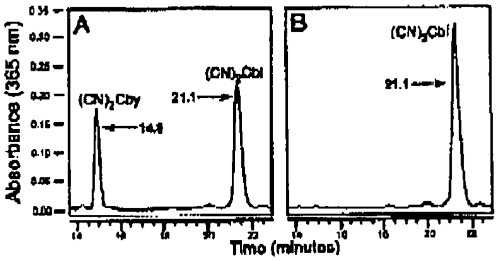


Fig. 5. HPLC analysis of the CbiZ reaction. Chromatograms of components of reaction mixture monitored at 365 nm (*A*) and the heat-inactivated control (*B*). Numbers represent times (in minutes) of elution after injection.

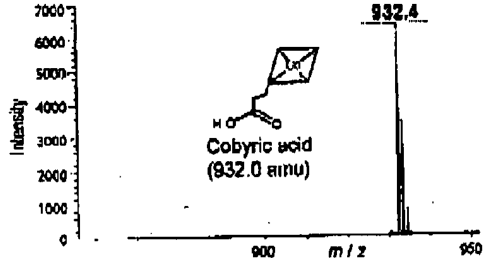


Fig. 6. MS analysis of the product of the CbiZ reaction. Shown is the matrix-assisted laser desorption/ionization-time-of-flight MS analysis of the HPLC-purified product of the CbiZ reaction. The signal with the m/z value of 932.4 was consistent with the molecular mass of Cby (without ligands) where $z = +1$. No significant signals were detected above an m/z value of 950.

inhibitory effects of the CN anion, substrate levels of KCN were added to the reaction mixture. KCN did not significantly affect the specific activity of the protein (data not shown). HPLC-purified reaction product was analyzed by MS. The signal with an $m/z = 932.4$ signal that was consistent with the expected molecular mass of Cby without any ligands (932.0 atomic mass units; Fig. 6). A bioassay confirmed the presence of Cby in the HPLC-purified peak (data not shown). These results confirmed that Cby was a product of the CbiZ reaction; i.e., that CbiZ had AdoCbi amidohydrolase enzyme activity.

Discussion

Archaea and Bacteria Salvage Cbi Via Two Distinct Pathways. The genetic and biochemical evidence reported here and elsewhere (20) supports the conclusion that prokaryotes have evolved at least two distinct pathways for salvaging the precursor Cbi from the environment. Information currently available from genome databases suggests that these Cbi-salvaging pathways evolved and remained segregated in separate domains of life; i.e., in the Archaea or the Bacteria. Both pathways accomplish the same goal, which is to convert AdoCbi to AdoCbi-P, a true intermediate of the *de novo* biosynthetic pathway (8, 9). The differences between the bacterial and archaeal Cbi salvaging pathways are illustrated in Figs. 1 and 7. The chief difference is the point of entry for AdoCbi. In this model, we assumed that archaea convert Cbi to AdoCbi, however, the identity of the ATP:co(I)rrinoid adenosyltransferase in archaea has yet to be established experimentally. At this point, it is unclear whether the substrate for CbiZ needs to be adenosylated in the cell. Under the *in vitro* conditions described in this paper, it is clear that $(CN)_2Cbi$ can be used as a substrate by CbiZ and the activity is only reduced 3-fold. Further characterization of the CbiZ protein may provide insight to when the corrin ring is adenosylated during salvaging in archaea.

Cobamide-producing bacteria evolved a conserved multifunctional enzyme that can use AdoCbi as substrate and convert it to AdoCbi-P (CobU in *S. enterica*) in a single catalytic step (Fig. 1). Archaea, on the other hand, convert AdoCbi to AdoCbi-P in two steps. First, the amidohydrolase activity of CbiZ cleaves off the aminopropanol moiety of AdoCbi yielding AdoCby; second, AdoCby is converted to AdoCbi-P by the action of the AdoCbi-P synthase (CbiB) enzyme (Fig. 7). Results from nutritional analysis of *cbiP* and *cbiB* mutants of *Halobacterium* sp. strain NRC-1 and complementation studies of *S. enterica* mutants unable to salvage Cbi using archaeal genes (20), strongly support the Cbi salvaging pathway delineated in Fig. 7 for archaea. Here and elsewhere (8, 19, 20), we have shown the existence of this pathway in euryarchaeotes. Whether this pathway is present in other archaea needs to be investigated. All available archaeal genome sequences contain orthologs of CbiZ and other proteins known or predicted to be required for the salvaging of Cbi (CobADSTY and CbiB), making it likely that the CbiZ-dependent Cbi-salvaging pathway is conserved among all archaea.

Although the bacterial CobU and the archaeal CbiZ enzymes are both used by cells to salvage Cbi, the enzymes share no sequence similarity, and in fact CbiZ does not share homology to any previously characterized proteins and contains no obvious motifs. The two enzymes, however, use AdoCbi as substrate. It will be interesting to see what, if any, structural similarities exist between how the two enzymes bind AdoCbi. No evidence for an ATP requirement was obtained. The structural analysis of the CbiZ protein warrants future study.

Although archaea and bacteria seem to have separate pathways for the salvaging of Cbi, putative orthologs of the CbiZ protein exist in *Bacillus halodurans* (30) and *Bacillus subtilis* (31). In both organisms, the CbiZ protein appears to be fused to the C terminus of the BtuD protein, the ATP-binding component of the B_{12} ABC-transport system. Because the CbiZ enzyme appears to be primarily a Cbi-salvaging function in archaea, a close association with the transport system should not be surprising. What is intriguing, however, is the role that CbiZ may be playing in *B. halodurans*, especially because this bacterium already has a putative ortholog to the bacterial ATP:AdoCbi kinase enzyme used by B_{12}-producing bacteria to salvage Cbi. In the case of *B. subtilis*, the role of CbiZ is even more obscure because this

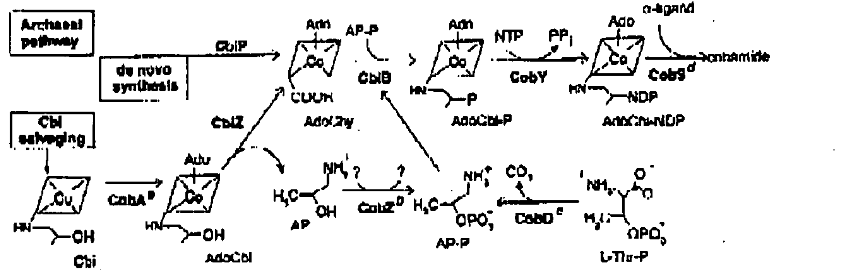


Fig. 7. Late steps of cobamide biosynthesis in archaea. Intermediates are indicated below structures. CobY, NTP:AdoCbi-P nucleotidyltransferase. *a*, The putative archaeal ortholog to the bacterial CobA protein. *b*, CobZ is the archaeal nonorthologous replacement of the bacterial CobC protein (C. L. Zayas, J.D.W., and J.C.E.-S., unpublished results). *c*, The archaeal CobD is the ortholog to the bacterial CobD protein (C. L. Zayas, J.D.W., and J.C.E.-S., unpublished results). *d*, The archaeal CobS is the ortholog to the bacterial CobS protein (32).

bacterium lacks any of the B_{12} biosynthetic enzymes. In the absence of other Cob enzymes, Cbi salvaging would be unlikely. CbiZ may be playing a different role in these bacteria.

Why Are Two Cbi-Salvaging Pathways Involved? It is unclear what selective pressures directed the evolution of two different Cbi-salvaging pathways in prokaryotes, and why these pathways were segregated to either bacteria or archaea. One could speculate that prokaryotes that constitutively express genes encoding *de novo* biosynthetic enzymes might have evolved the amidohydrolase route, in response to the stability of the intermediate to which AdoCbi would be converted. It is possible that AdoCbi-P (the result of a hypothetical kinase) would not be stable enough for the next enzyme of the pathway to use it as substrate, whereas AdoCby, the product of CbiZ, could be.

Prokaryotes that conditionally express the corrin ring biosynthetic functions, but constitutively express functions required for the assembly of the nucleotide loop (e.g., *S. enterica*) face a different problem. Under some conditions, these organisms could find Cbi in the environment, but whether such growth conditions were not conducive for the expression of the AdoCbi-P synthase (CbiB) enzyme (e.g., presence of oxygen), the organism would be unable to make coenzyme B_{12} from AdoCbi, whether salvaging of the latter depended on the activity of an amidohydrolase enzyme. This problem would be circumvented through the evolution of a kinase enzyme to convert AdoCbi to AdoCbi-P by direct phosphorylation. The fact that such AdoCbi kinase activity has only been found in a protein that also has the next catalytic activity (GTP:AdoCbi-P guanylyl transferase) of the pathway suggests that AdoCbi P may be too unstable to be released from the enzyme.

We thank P. Renz for his gift of Cby and G. Gottschalk for his gift of *M. mazei* chromosomal DNA. This work was supported by National Institutes of Health Grant GM40313 (to J.C.E.-S.). J.D.W. was supported in part by the Ira L. Baldwin and Jerome Stefaniak Predoctoral Fellowships.

1. Roth, J. R., Lawrence, J. G. & Bobik, T. A. (1996) *Annu. Rev. Microbiol.* 50, 137–181.
2. Scott, A. I. & Roessner, C. A. (2002) *Biochem. Soc. Trans.* 30, 613–620.
3. Warren, M. J., Raux, E., Schubert, H. L. & Escalante-Semerena, J. C. (2002) *Nat. Prod. Rep.* 19, 390–412.
4. Rondon, M. R., Trzebiatowski, J. R. & Escalante-Semerena, J. C. (1997) *Prog. Nucleic Acid Res. Mol. Biol.* 56, 347–384.
5. Raux, E., Lanois, A., Warren, M. J., Rambach, A. & Thermes, C. (1998) *Biochem. J.* 335, 159–166.
6. Santander, P. J., Roessner, C. A., Stolowich, N., Holderman, M. T. & Scott, A. I. (1997) *Chem. Biol.* 4, 659–666.
7. Blanche, F., Cameron, B., Crouzet, J., Debussche, L., Thibaut, D., Vuilhorgne, M., Leeper, F. J. & Battersby, A. R. (1995) *Angew. Chem. Int. Ed. Engl.* 34, 383–411.
8. Thomas, M. G. & Escalante-Semerena, J. C. (2000) *J. Bacteriol.* 182, 4227–4233.
9. Brushaber, K. R., O'Toole, G. A. & Escalante-Semerena, J. C. (1998) *J. Biol. Chem.* 273, 2684–2691.
10. Fonseca, M. V. & Escalante-Semerena, J. C. (2000) *J. Bacteriol.* 182, 4304–4309.
11. Fonseca, M. V. & Escalante-Semerena, J. C. (2001) *J. Biol. Chem.* 276, 32101–32108.
12. Fonseca, M. V., Buan, N. R., Horswill, A. R., Rayment, I. & Escalante-Semerena, J. C. (2002) *J. Biol. Chem.* 277, 33127–33131.
13. Bauer, C. B., Fonseca, M. V., Holden, H. M., Thoden, J. B., Thompson, T. B., Escalante-Semerena, J. C. & Rayment, I. (2001) *Biochemistry* 40, 361–374.
14. O'Toole, G. A. & Escalante-Semerena, J. C. (1995) *J. Biol. Chem.* 270, 23560–23569.
15. Thomas, M. G., Thompson, T. B., Rayment, I. & Escalante-Semerena, J. C. (2000) *J. Biol. Chem.* 275, 27376–27386.
16. Thompson, T., Thomas, M. G., Escalante-Semerena, J. C. & Rayment, I. (1999) *Biochemistry* 38, 12995–13005.
17. Thompson, T. B., Thomas, M. G., Escalante-Semerena, J. C. & Rayment, I. (1998) *Biochemistry* 37, 7686–7695.
18. Blanche, F., Debussche, L., Famechon, A., Thibaut, D., Cameron, B. & Crouzet, J. (1991) *J. Bacteriol.* 173, 6052–6057.
19. Woodson, J. D., Peck, R. F., Krebs, M. P. & Escalante-Semerena, J. C. (2003) *J. Bacteriol.* 185, 311–316.
20. Woodson, J. D., Zayas, C. L. & Escalante-Semerena, J. C. (2003) *J. Bacteriol.* 185, 7193–7201.
21. Tsai, S. P., Hartin, R. J. & Ryu, J. (1989) *J. Gen. Microbiol.* 135, 2561–2567.
22. Rondon, M. R. & Escalante-Semerena, J. C. (1992) *J. Bacteriol.* 174, 2267–2272.
23. Laemmli, U. K. (1970) *Nature* 227, 680–685.
24. Sauer, B. (1991) in *Current Protocols in Molecular Biology*, eds. Ausubel, F. A., Brent, R., Kingston, R. E., Moore, D. D., Seidman, J. G., Smith, J. A. & Struhl, K. (Wiley, New York), Vol. 1, pp. 10.6.1–10.6.8.
25. Blanche, F., Thibaut, D., Couder, M. & Muller, J. C. (1990) *Anal. Biochem.* 189, 24–29.
26. Deppenmeier, U., Johann, A., Hartsch, T., Merkl, R., Schmitz, R. A., Martinez-Arias, R., Henne, A., Wiezer, A., Baumer, S., Jacobi, C., *et al.* (2002) *J. Mol. Microbiol. Biotechnol.* 4, 453–461.
27. Drennan, C. L., Huang, S., Drummond, J. T., Matthews, R. G. & Ludwig, M. L. (1994) *Science* 266, 1669–1674.
28. O'Toole, G. A. (1994) Ph.D. thesis (Univ. of Wisconsin, Madison, WI).
29. Grabau, C. & Roth, J. R. (1992) *J. Bacteriol.* 174, 2138–2144.
30. Takami, H., Nakasone, K., Takaki, Y., Maeno, G., Sasaki, R., Masui, N., Fuji, F., Hirama, C., Nakamura, Y., Ogasawara, N., *et al.* (2000) *Nucleic Acids Res.* 28, 4317–4331.
31. Kunst, F., Ogasawara, N., Moszer, I., Albertini, A. M., Alloni, G., Azevedo, V., Bertero, M. G., Bessieres, P., Bolotin, A., Borchert, S., *et al.* (1997) *Nature* 390, 249–256.
32. Maggio-Hall, L. A. (2001) in *Department of Bacteriology* (Univ. of Wisconsin, Madison, WI), p. 155.

Vol. 186, No. 2

JOURNAL OF BACTERIOLOGY, Jan. 2004, p. 454–462
0021-9193/04/$08.00+0 DOI: 10.1128/JB.186.2.454–462.2004
Copyright © 2004, American Society for Microbiology. All Rights Reserved.

The *acnD* Genes of *Shewenella oneidensis* and *Vibrio cholerae* Encode a New Fe/S-Dependent 2-Methylcitrate Dehydratase Enzyme That Requires *prpF* Function In Vivo

Tracey L. Grimek and Jorge C. Escalante Semerena*

Department of Bacteriology, University of Wisconsin—Madison, Madison, Wisconsin

Received 24 July 2003/Accepted 10 October 2003

The propionate utilization operons of several bacteria differ from each other in the occurrence of two genes, *acnD* and *prpF*, in place of or in addition to the *prpD* gene encoding an Fe/S-independent 2-methylcitrate dehydratase enzyme. We cloned the *acnD* and *prpF* genes from two organisms, *Shewanella oneidensis* and *Vibrio cholerae*, and found that, together, the AcnD and PrpF proteins restored the ability of a *prpD* mutant strain of *Salmonella enterica* to grow on propionate as a source of carbon and energy. However, neither *acnD* nor *prpF* alone was able to substitute for *prpD*. The AcnD and PrpF proteins were isolated and biochemically analyzed. The AcnD protein required reconstitution of an Fe/S cluster for activity. All detectable AcnD activity was lost after incubation with iron-chelating agents, and no AcnD activity was observed after attempted reconstitution without iron. Nuclear magnetic resonance spectroscopy and in vitro activity assay data showed that AcnD dehydrated 2-methylcitrate and citrate to 2-methyl-*cis*-aconitate and *cis*-aconitate, respectively; AcnD also hydrated *cis*-aconitate. However, 2-methylisocitrate and isocitrate were not substrates for AcnD, indicating that AcnD only catalyzes the first half of the aconitase-like dehydration reactions. No aconitase-like activity was found for PrpF. It is hypothesized that, in vivo, PrpF is an accessory protein required to prevent oxidative damage of the Fe/S center of active AcnD enzyme or that it may be involved in synthesis or repair of the Fe/S cluster present in AcnD.

First demonstrated in *Yarrowia lipolytica* and several other filamentous fungi and yeast species (26, 32, 33), the 2-methylcitrate (2-MC) cycle was subsequently shown to occur in the bacteria *Salmonella enterica* serovar Typhimurium and *Escherichia coli* (20, 35). Since the finding that prokaryotes can also utilize the 2-MC cycle as their route of propionate catabolism, the 2-MC pathway was also demonstrated in other gram-negative bacteria such as *Ralstonia eutropha* (8) and *Burkholderia sacchari* (7) and in the gram-positive bacterium *Corynebacterium glutamicum* (10). Sequence analysis of the completed genomes of several other bacteria indicates that the 2-MC cycle may be widespread among bacteria; *Vibrio cholerae*, *Shewanella oneidensis*, *Neisseria* spp., and several *Pseudomonas* species contain propionate utilization (*prp*) operons (8, 17) (Fig. 1).

The *prp* operons of *S. enterica* and *E. coli* comprise four genes encoding structural proteins of the 2-MC cycle which have been characterized (Fig. 1A) as follows: *prpB* encodes 2-methylisocitrate (2-MIC) lyase (13, 14, 17), *prpC* encodes 2-MC synthase (17, 18, 35), *prpD* encodes 2-MC dehydratase (9, 17), and *prpE* encodes propionyl-coenzyme A synthetase (19). However, other *prp* operons have a gene organization that differs greatly from that of these two enterics (8, 17) (Fig. 1). Figure 1D shows an operon structure that contains two genes, *acnD* (17) and *prpF*, instead of *prpD*. In *S. enterica*, *prpD* encodes an Fe/S-independent 2-MC dehydratase that generates 2-methyl-*cis*-aconitate (2-MCA) from 2-MC but will not hydrate 2-MCA into 2-MIC (17). The hydration of 2-MCA is catalyzed by either aconitase AcnA or AcnB (Fig. 2) (17). All currently sequenced *prp* operons that contain an *acnD* ortholog also contain *prpF* and vice versa. The latter is an ortholog of open reading frame 5 (ORF5) of *R. eutropha* (8) and *E. coli ybhH*. The only work on *acnD* and *prpF* reported to date was performed in *R. eutropha* (8). The *R. eutropha prp* operon, represented in Fig. 1C, contains both the *acnD* and *prpF* genes and *prpD*. Brämer et al. recently reported that, in *R. eutropha* HF39, *prpD* was not required for a functional 2-MC cycle, but *acnD* and *prpF* (*acnM* and ORF5 in *R. eutropha*, respectively) functions were needed. However, efforts to demonstrate the conversion of 2-MC to 2-MIC by AcnM-enriched crude cell extracts were inconclusive (8).

In this paper we demonstrate that the *acnD* gene encodes an Fe/S-dependent 2-MC dehydratase enzyme that requires the *prpF* gene product to function in vivo. The *acnD* and *prpF* genes from *V. cholerae* and *S. oneidensis* were cloned independently or together and were used to compensate for the lack of the Fe/S-independent 2-MC dehydratase (PrpD) enzyme in *S. enterica* during growth on propionate. The AcnD and PrpF proteins were isolated. AcnD purified in the presence of air was inactive but was reactivated by protocols reported for the reactivation of aconitase (22). Reactivated AcnD had 2-MC dehydratase activity but no measurable 2-MIC dehydratase activity. Even though aconitase-like activity was not observed for PrpF in vitro, PrpF was required for the conversion of 2-MC into 2-MCA in vivo. Possible roles for the PrpF protein are discussed.

---

* Corresponding author. Mailing address: Department of Bacteriology, University of Wisconsin, 264 Enzyme Institute, 1710 University Ave., Madison, WI 53726-4087. Phone: (608) 262-7379. Fax: (608) 265-7909. E-mail: escalante@bact.wisc.edu.

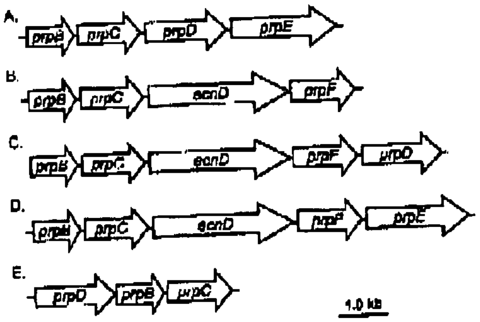


FIG. 1. Structural variations on *prp* operons of several bacteria. (A) *S. enterica* serovar Typhimurium and *E. coli* (*E. coli* contains a 439-nt putative stem-loop region between *prpB* and *prpC*). (B) *R. eutropha* CH34, *S. oneidensis* MR-1, *Neisseria meningitidis*, *Neisseria gonorrhoeae* (the *Neisseria* spp. contain a 789-nt ORF of unknown function between *prpC* and *acnD*, which has been designated *yfiA*, and an ORF of 1,300 nt following *prpF* designated *ackA* that shows sequence similarity to propionate kinase [*tdcD*] of *E. coli*), and *B. sacchari* IPT101 (*B. sacchari* contains a putative ORF of 366 nt between *acnD* and *prpF* and two putative ORFs of 372 nt and 198 nt following *prpF*). (C) *R. eutropha* HF39, *Bordetella pertussis*, *Pseudomonas aeruginosa*, and *Pseudomonas putida* KT2440. (D) *V. cholerae*. (E) *prpDBC2* operon of *C. glutamicum*. Putative regulators have been excluded from Fig. 1C and D, and the spaces between ORFs are not drawn to scale.

MATERIALS AND METHODS

Chemicals and culture media. Cultures were maintained in Luria-Bertani (LB) broth and solid media. No-carbon E medium supplemented with $MgSO_4$ (1 mM) and methionine (0.5 mM) was used as minimal medium (5, 11). Propionate and pyruvate were used at concentrations of 30 mM. Antibiotic concentrations in rich media were as follows (in µg/ml): ampicillin, 100; kanamycin, 25 (for plasmids, 50 µg/ml); tetracycline, 15; and chloramphenicol, 20. Bacterial strains harboring plasmids were grown in minimal media containing ampicillin (50 µg/ml) and kanamycin (30 µg/ml). Synthetic 2-MC was purchased from C/D/N Isotopes (Pointe-Claire, Quebec, Canada) as a mixture of stereoisomers; [2-^{13}C]propionate, 100% deuterium oxide (D_2O), and tetramethylsilane were purchased from Cambridge Isotope Labs (Andover, Mass.). The 2-MIC was a gift from W. W. Cleland. All other chemicals were purchased from Sigma (St. Louis, Mo.) unless otherwise stated. A list of the strains and plasmids used and their genotypes is provided in Table 1.

Recombinant DNA techniques. Restriction and modification enzymes were purchased from Promega (Madison, Wis.) unless otherwise stated and were used according to the manufacturer's specifications. All DNA manipulations were performed in *E. coli* strain DH5α/F'. Plasmids were transformed into *S. enterica* strains by a quick-electroporation method as follows. Strains were grown to approximately mid-exponential phase, and cells in 1.0 ml of culture were pelleted at 10,000 × *g* in a Microfuge 18 centrifuge (Beckman Coulter). Cells were washed once with 100 µl of cold H_2O, resuspended in 100 µl of cold H_2O, and allowed to equilibrate on ice for 5 min. Plasmids were electroporated into the competent cells with a Bio-Rad Gene Pulser (Hercules, Calif.), according to the manufacturer's recommendations.

Construction of plasmids. Plasmids containing *S. oneidensis* or *V. cholerae* genes were constructed by PCR amplification of genomic DNA of *S. oneidensis* MR-1 (a gift from D. Saffarini, University of Wisconsin—Milwaukee) or *V. cholerae* N16961 (a gift from Ron Taylor, Dartmouth Medical School). PCRs typically contained the following in a 100-µl reaction mixture: 1.5 ng of genomic DNA, 50 pmol of each primer (IDT DNA, Coralville, Iowa), and deoxynucleoside triphosphate and HiFi DNA polymerase (Novagen, Madison, Wis.), each at a concentration of 0.2 µM, according to manufacturer's instructions. Reactions were performed under the following conditions: 35 cycles at 95°C for 30 s, at 50°C for 30 s, and at 72°C for 1 min per kb of target DNA. The PCR fragment was purified with a QIAquick PCR purification kit (QIAGEN, Chatsworth, Calif.). The methods of constructing the plasmids used are outlined in Table 2. The primer sequences used in plasmid constructions are available upon request.

Sequence verification of plasmid constructs. All resulting plasmid constructs were sequenced to verify that no mutations were introduced into the genes of interest. PCR sequencing reactions were prepared with Big Dye (Biotechnology Center, University of Wisconsin—Madison). Reactions were purified by means of the CleanSEQ reaction clean-up protocol of Agencourt Bioscience Corporation (Beverly, Mass.) and sequenced at the Biotechnology Center. Searches for sequence similarity were performed by using the BLAST algorithm (1). Protein sequence alignments were conducted with the ClustalW multiple alignment tool (36).

Complementation analysis. Plasmids were introduced into *S. enterica* strains as described above. The resulting strains were grown overnight in LB broth containing appropriate antibiotics. Four microliters of each overnight culture was used to inoculate 200 µl of fresh no-carbon E minimal medium supplemented with propionate (30 mM) and glycerol (1 mM) or acetate (30 mM), the appropriate antibiotic, and various amounts of L-(+)-arabinose (0, 100, or 500 µM). Medium was placed into the wells of a 96-well Falcon (Becton Dickinson, Franklin Lakes, N.J.) microtiter dish, and the density of the cultures was monitored at 650 nm with a SpectraMAX Plus high-throughput spectrophotometer (Molecular Devices, Sunnyvale, Calif.). The plate chamber in the spectrometer was maintained at 37°C. Absorbance measurements were taken every 15 min for 72 h with agitation (for 780 s) between reads.

Anaerobic growth analysis. For anaerobic growth experiments, LB plates containing 10 mM sodium tetrathionate as a terminal electron acceptor (27) were patched with the strains of interest and introduced into an anoxic environment with Gas-Pak jars (Bethesda Research Laboratories, Gaithersburg, Md.). Plates were incubated at 37°C for 24 h and then transferred into an anaerobic chamber, where they were replica printed onto minimal propionate medium plates supplemented with 10 mM tetrathionate and various concentrations of arabinose. Plates were incubated anoxically for 3 days at 37°C and growth was assessed.

Purification of *Shewanella* AcnD and PrpF proteins. Plasmids pPRP152 (*S. oneidensis prpF*$^+$) and pPRP156 (*S. oneidensis acnD*$^+$) were introduced into *E. coli* BL21(λDE3) by the $CaCl_2$ heat shock method described elsewhere (30). Cells (20 ml of an overnight culture) were inoculated into 2 liters of LB broth supplemented with 100 µg of ampicillin/ml and grown with shaking at 37°C. Cells were grown to an A_{600} of approximately 0.6, and the overproduction of proteins was induced with 0.3 mM isopropyl-β-D-thiogalactopyranoside (IPTG). The cultures were harvested 3 h after induction for 10 min at 4°C and 11,700 × *g*. The cell pellets were frozen at −70°C for later use.

Cell pellets were resuspended in 25 ml of 20 mM (pH 7.5) 4-(2-hydroxyethyl)-

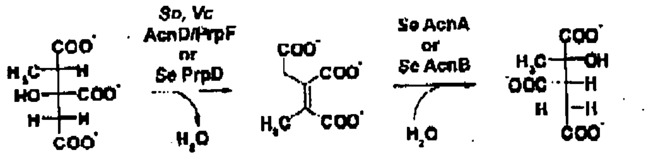


FIG. 2. Conversion of 2-MC to 2-MIC in bacteria. 2-MC and 2-MIC are drawn as Fischer projections; stereochemistry is based on intermediates of the *E. coli* 2-MC cycle (9). *Se*, *S. enterica*; *Vc*, *V. cholerae*; *So*, *S. oneidensis*.

TABLE 1. Strains and plasmids used in this study[a]

Strain or plasmid	Genotype	Reference or source
E. coli strains		
BL21(λDE3)	F$^-$ *ompT hsdSB*(r_B^- m_B^-) *dcm gal* λ(DE3)	New England Biolabs
DH5α/F′	F′/*endA1 hsdR17*(r_K^- m_K^+) *supE44 thi-1 recA1 gyrA* (Nal^r) *relA1* Δ(*lacZYA-argF*)U169 *deoR* [φ80*dlac*Δ(*lacZ*)*M15*]	New England Biolabs
JE4570	BL21(λDE3)/pPRP62 (*S. enterica prpC*$^+$ in pET15b *bla*$^+$)	
JE4744	BL21(λDE3)/pPRP67 (*S. enterica prpD*$^+$ in pET15b *bla*$^+$)	
S. enterica strains		
TR6583[b]	*metE205 ara-9*	K. Sanderson via J. Roth
Derivatives of TR6583[c]		
JE3056	*prpR121*::Tn*10*d(Tc)	18
JE3907	*prpC167 zai-6386*::Tn*10*d(Tc)[d]	18
JE3909	*prpD169 zai-6386*::Tn*10*d(Tc)	18
JE3946	*prpB195 zai-6386*::Tn*10*d(Tc)	18
JE5993	*acnA2*::*cat*$^+$ *acnB3*::*kan*$^+$	17
JE6501	JE5993/pBAD30 (*bla*$^+$)	
JE6502	JE5993/pACN9 (*S. enterica acnA*$^+$ in pBAD30 *bla*$^+$)	
JE6503	JE5993/pACN10 (*S. enterica acnB*$^+$ in pBAD30 *bla*$^+$)	
JE6504	JE5993/pPRP121 (*V. cholerae acnD*$^+$ *prpF*$^+$ in pBAD30 *bla*$^+$)	
JE6506	JE5993/pPRP123 (*V. cholerae acnD*$^+$ *prpF*$^+$ in pBAD30 *bla*$^+$)	
JE7235	JE5993/pPRP138 (*S. oneidensis acnD*$^+$ in pBAD30 *bla*$^+$)	
JE7236	JE5993/pPRP140 (*S. oneidensis acnD*$^+$ *prpF*$^+$ in pBAD30 *bla*$^+$)	
JE7357	JE5993/pPRP163 (*E. coli ybhJ*$^+$ in pBAD30 *bla*$^+$)	
JE7590	JE5993/pPRP166 (*E. coli ybhHI*$^+$ in pBAD30 *bla*$^+$)	
Plasmids		
pBAD30	P_{araBAD} expression vector, *bla*$^+$	16
pACN9	*S. enterica acnA*$^+$ in pBAD30 *bla*$^+$	17
pACN10	*S. enterica acnB*$^+$ in pBAD30 *bla*$^+$	17
pPRP12-5.4	*S. enterica prpBCD*$^+$ in pSU38 *kan*$^+$	18
pPRP21	*S. enterica prpB*$^+$ in pBAD30 *bla*$^+$	18
pPRP36	*S. enterica prpD*$^+$ in pBAD30 *bla*$^+$	18
pPRP35	*S. enterica prpC*$^+$ in pBAD30 *bla*$^+$	18
pPRP62	*S. enterica prpC*$^+$ in pET15b *bla*$^+$	20
pPRP67	*S. enterica prpD*$^+$ in pET15b *bla*$^+$	17
pPRP121	*V. cholerae acnD*$^+$ in pBAD30 *bla*$^+$	
pPRP123	*V. cholerae acnD*$^+$ *prpF*$^+$ in pBAD30 *bla*$^+$	
pPRP138	*S. oneidensis acnD*$^+$ in pBAD30 *bla*$^+$	
pPRP140	*S. oneidensis acnD*$^+$ *prpF*$^+$ in pBAD30 *bla*$^+$	
pPRP141	*S. oneidensis prpB*$^+$ *prpC*$^+$ *acnD*$^+$ *prpF*$^+$ in pBAD30 *bla*$^+$	
pPRP149	*S. oneidensis prpB*$^+$ in pBAD18Kan *kan*$^+$	
pPRP150	*S. oneidensis prpC*$^+$ in pBAD18Kan *kan*$^+$	
pPRP151	*S. oneidensis prpB*$^+$ *prpC*$^+$ in pBAD18Kan *kan*$^+$	
pPRP152	*S. oneidensis acnD*$^+$ in pTYB12 *bla*$^+$	
pPRP153	*S. oneidensis prpF*$^+$ in pBAD18Kan *kan*$^+$	
pPRP154	*E. coli ybhH*$^+$ in pBAD18Kan *kan*$^+$	
pPRP155	*V. cholerae prpF*$^+$ in pBAD18Kan *kan*$^+$	
pPRP156	*S. oneidensis prpF*$^+$ in pTYB12 *bla*$^+$	
pPRP163	*E. coli ybhJ*$^+$ in pBAD30 *bla*$^+$	
pPRP166	*E. coli ybhHI*$^+$ in pBAD30 *bla*$^+$	

[a] All *S. enterica* strains are derivatives of the *S. enterica* serovar Typhimurium LT2 strain. Unless otherwise stated, strains and plasmids were constructed during the course of this work.

[b] Formerly SA2079.

[c] See Table 3 for additional strains (and their genotypes) used in this work.

[d] Tn*10*d(Tc) is an abbreviation of Tn*10*DEL*16*DEL*17* (37)

1-piperazineethanesulfonic acid (buffer A) (HEPES; Fisher Biotech, Itasca, Ill.), containing 100 mM KCl, 0.1% (vol/vol) Triton X-100, and 0.1 mM EDTA. The cell suspensions were broken at 10^4 kPa in a chilled French pressure cell. Cell debris was removed by centrifugation at 31,000 × *g* for 30 min at 4°C. Crude cell extracts were filtered through a 0.2-μm-pore-size filter and passed through a 5-ml column of chitin beads (New England Biolabs, Beverly, Mass.). The column was washed with buffer A according to the manufacturer's instructions and then quickly washed with 15 ml of buffer A containing 30 mM 1,4-dithio-DL-threitol (DTT; Promega, Madison, Wis.). The column flow was stopped, and the column was kept at 4°C for 46 h. Purified, untagged proteins were eluted off the column with buffer A containing 50 mM DTT and were visualized with sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) (23). Fractions containing protein were pooled and dialyzed overnight into 20 mM HEPES (pH 7.5), 100 mM KCl, 0.1 mM EDTA, and 5 mM DTT (buffer B). After dialysis for 2 h, the buffer was changed to buffer B lacking EDTA. The remaining part of the dialysis period was performed with buffer B containing 5% (vol/vol) glycerol. The protein was flash-frozen in liquid nitrogen and stored at −80°C.

Reactivation of purified proteins. Purified AcnD was reactivated according to the method of Kennedy and Beinert (22). All reagents were prepared under strict anaerobic conditions (3, 15). Purified H_6AcnA (17) was reactivated by the same procedure and was used as a positive control in the aconitase assays.

***In vitro* aconitase assays.** Aconitase assays were performed as described (17). Reaction mixtures contained 90 mM Tris (Tris-HCl) buffer (pH 8.0) and 20 mM citrate, isocitrate, 2-MC, 2-MIC, or 2 mM *cis*-aconitate in a 1-ml quartz cuvette. Reactivated enzyme was transferred to the cuvette with a 25-μl Hamilton syringe. No detectable loss of activity was observed during the duration of the assays (2 h). Reactions were monitored for 10 min at 240 nm in a Perkin-Elmer Lambda 40 spectrophotometer (Norwalk, Conn.) equipped with a circulating

TABLE 2. Construction of plasmids used in this work

Plasmid name	Vector backbone	Gene(s) cloned	Method of construction
pPRP117	pDONR201[a]	*V. cholerae acnD*	Gateway BP reaction[b]
pPRP121	pBAD30	*V. cholerae acnD*	Gateway LR reaction
pPRP123	pBAD30	*V. cholerae acnD prpF*	*Kpn*I, *Xba*I cloning
pPRP136	pDONR201	*S. oneidensis acnD*[c]	Gateway BP reaction
pPRP138	pBAD30	*S. oneidensis acnD*	Gateway LR reaction
pPRP137	pDONR201	*S. oneidensis acnD prpF*	Gateway BP reaction
pPRP140	pBAD30	*S. oneidensis acnD prpF*	Gateway LR reaction
pPRP141	pBAD30	*S. oneidensis prpB prpC acnD prpF*	*Sac*I, *Xba*I cloning
pPRP149	pBAD18-Kan	*S. oneidensis prpB*	*Sac*I, *Xba*I cloning
pPRP150	pBAD18-Kan	*S. oneidensis prpC*	*Sac*I, *Xba*I cloning
pPRP151	pBAD18-Kan	*S. oneidensis prpB prpC*	*Sac*I, *Xba*I cloning
pPRP152	pTYB12	*S. oneidensis prpF*	*Bsm*I, *Xho*I cloning
pPRP153	pBAD18-Kan	*S. oneidensis prpF*	*Eco*RI, *Xba*I cloning
pPRP154	pBAD18-Kan	*E. coli*[d] *ybhH*	*Sac*I, *Xba*I cloning
pPRP155	pBAD18-Kan	*V. cholerae prpF*	*Kpn*I, *Xba*I cloning
pPRP156	pTYB12	*S. oneidensis acnD*	*Bsm*I, *Xho*I cloning
pPRP163	pBAD30	*E. coli ybhJ*	*Sac*I, *Xba*I cloning
pPRP166	pBAD30	*E. coli ybhH ybhI ybhJ*	*Sac*I, *Xba*I cloning

[a] pDONR201 was purchased from Invitrogen, Carlsbad, Calif.
[b] Gateway BP and LR reactions are methods developed by Life Technologies, Invitrogen Corporation.
[c] All *S. oneidensis* organisms are strain MR-1.
[d] All *E. coli* organisms are strain MG1655.

water bath, which maintained the temperature at 37°C. Specific activities were reported in micromoles per minute per milligram of protein and calculated from the extinction coefficients of 3,600 M^{-1} cm^{-1} for the cis-aconitate (21) and 4,500 M^{-1} cm^{-1} for the 2-methyl-*cis*-aconitate (2).

Requirement of an Fe/S cluster for AcnD activity. To determine if iron was required for AcnD activity, AcnD was reactivated according to the protocol of Kennedy and Beinert (22) with iron excluded from the reactivation mixture. To determine the magnitude of inactivation of AcnD in the presence of iron-chelating agents, reactivated AcnD was incubated with EDTA and ferricyanide in the molar ratios of 1:50:20 (enzyme:EDTA:ferricyanide) as described by Kennedy and Beinert (22). Protein activity was assayed as described above.

H_6PrpC, H_6PrpB, and PrpE proteins. The *S. enterica* H_6PrpC and H_6PrpB proteins were overproduced and purified as previously described (17, 20). The propionyl-coenzyme A synthetase (PrpE) enzyme was provided by V. J. Starai.

^{13}C-NMR spectroscopy. Peak assignments were compared to those previously reported (20). Glycerol was present in all protein samples and observed in the spectra due to the natural abundance of ^{13}C (17). ^{13}C-nuclear magnetic resonance (NMR) spectra were obtained at the National Magnetic Resonance Facility at the University of Wisconsin—Madison by means of a Bruker Instruments DMX-400 Avance console with a 9.4 T wide-bore magnet at 100.6 MHz.

In vitro enzymatic synthesis of [2-^{13}C]MC. [2-^{13}C]MC was generated in vitro (17) in 0.5-ml reaction mixtures that contained potassium phosphate buffer (pH 7.5; 50 mM), ATP (2.5 mM), $MgCl_2$ (5 mM), coenzyme A (2.5 mM), [2-^{13}C]propionate (2.5 mM), oxaloacetate (5 mM), and homogeneous PrpE and H_6PrpC proteins (75 µg each). To each sample, 0.1 ml of 100% D_2O was added, and the reactions were transferred to 5-mm NMR tubes (Wilmad Glass, Buena, N.J.). A sealed tetramethylsilane capillary was added as an external reference.

Conversion of [2-^{13}C]MC to [2-^{13}C]MCA. [2-^{13}C]MC (~2.5 mM) synthesized as described above was used as substrate for PrpF and reactivated AcnD or for H_6PrpD. Reaction mixtures (0.5 ml) contained [2-^{13}C]MC, H_6PrpD, PrpF, or reactivated AcnD (25 µg each) or PrpF and reactivated AcnD (25 µg each). The reaction mixtures were allowed to incubate for 1 h at 37°C and were prepared for ^{13}C-NMR analysis as described above.

Other procedures. Protein concentrations were determined from a standard curve generated with bovine serum albumin by the method of Bradford (6) with the Bio-Rad protein reagent. Proteins were separated by SDS–12% PAGE and stained with Coomassie blue (28). Novagen Perfect Protein markers (Madison, Wis.) were used as standards for SDS-PAGE.

RESULTS

***S. oneidensis* and *V. cholerae acnD* and *prpF* gene product functions restore growth of a *S. enterica prpD* mutant strain on propionate.** As pointed out above, *prp* loci from various prokaryotes contain two genes (*acnD* and *prpF*) in lieu of *prpD* (Fig. 1). The *acnD* and *prpF* genes from *S. oneidensis* and *V. cholerae* were cloned and used to determine whether they would compensate for the lack of PrpD function during growth of an *S. enterica prpD* mutant strain on propionate. The *acnD* and *prpF* genes were cloned into plasmids under the control of arabinose-inducible promoters either as a pair or individually on compatible plasmids. Plasmids carrying these genes were introduced into *S. enterica* strain JE3909 (*prpD*), and growth on propionate was assessed. Table 3 shows the doubling times of all strains tested. Both *S. oneidensis acnD$^+$prpF$^+$* and *V. cholerae acnD$^+$prpF$^+$* constructs complemented strain JE3909 (Fig. 3A). When the *acnD* and *prpF* genes were carried in separate plasmids, however, growth on propionate was observed only when both genes were present in the cell. In all cases when *S. oneidensis* and *V. cholerae* genes were mixed, strain JE3909 was able to grow on propionate (Fig. 3B). This result was not surprising considering that the *S. oneidensis* and *V. cholerae* AcnD and PrpF share 76 and 73% identity, respectively. These results indicate that both *acnD* and *prpF* are required to complement a *prpD* strain of *S. enterica*.

To determine whether PrpF function was required along with AcnD function to complement an *S. enterica prpD* mutant grown on a propionate medium anaerobically, these strains were replica printed onto a propionate medium containing tetrathionate as a terminal electron acceptor. Only the positive control (*prpD$^+$*) and strains containing both the *acnD$^+$* and *prpF$^+$* plasmids grew anaerobically on propionate. These data indicate that even under anaerobic conditions, both AcnD and PrpF functions are required to complement an *S. enterica prpD* mutant.

The possibility that the *E. coli ybhJ* and *ybhH* genes could restore the growth of strain JE3909 on propionate was also assessed. The *E. coli* YbhJ protein is an AcnA homolog that shares 22% identity (37% similarity) with *S. oneidensis* AcnD; the YbhH protein shares 34% identity (47% similarity) with *S.*

TABLE 3. Heterologous complementation of *S. enterica prp* mutant strains

Strain	Genotype[a]	Arabinose concentration[b] (µM)	Doubling time[c] (h ± SD)
JE4175	TR6583 (wild type)/pBAD30	500	7.3 ± 0.3
JE6107	JE3909/pPRP36 (*S. enterica prpD*$^+$)	500	7.6 ± 0.3
JE7280	JE3909/pPRP138 (*S. oneidensis acnD*$^+$), pPRP153 (*S. oneidensis prpF*$^+$)	500	7.9 ± 0.3
JE7232	JE3909/pPRP140 (*S. oneidensis acnD*$^+$ *prpF*$^+$)	500	8.0 ± 0.5
JE7227	JE3909/pPRP141 (*S. oneidensis prpD*$^+$ *prpC*$^+$ *acnD*$^+$ *prpF*$^+$)	500	9.2 ± 1.0
JE7282	JE3909/pPRP138 (*S. oneidensis acnD*$^+$), pPRP155 (*V. cholerae prpF*$^+$)	500	20.4 ± 1.2
JE7283	JE3909/pPRP121 (*V. cholerae acnD*$^+$), pPRP153 (*S. cholerae prpF*$^+$)	100	11.0 ± 0.5
JE7285	JE3909/pPRP121 (*V. cholerae acnD*$^+$), pPRP155 (*V. cholerae prpF*$^+$)	500	25.8 ± 2.0
JE7234	JE3909/pPRP123 (*V. cholerae acnD*$^+$ *prpF*$^+$)	100	13.0 ± 0.4
JE6105	JE3909/pBAD30	500	NG
JE7238	JE3056/pPRP12–5.4 (*S. enterica prpB*$^+$ *prpC*$^+$ *prpD*$^+$)	0	15.8 ± 0.5
JE7274	JE3056/pPRP151 (*S. oneidensis prpB*$^+$ *prpC*$^+$), pPRP36 (*S. oneidensis prpD*$^+$)	100	12.7 ± 0.3
JE7275	JE3056/pPRP151 (*S. oneidensis prpB*$^+$ *prpC*$^+$), pPRP140 (*S. oneidensis acnD*$^+$ *prpF*$^+$)	100	19.2 ± 0.4
JE7277	JE3056/pPRP151 (*S. oneidensis prpB*$^+$ *prpC*$^+$), pPRP123 (*V. cholerae acnD*$^+$ *prpF*$^+$)	100	14.7 ± 0.5
JE7239	JE3056/pPRP141 (*S. oneidensis prpB*$^+$ *prpC*$^+$ *acnD*$^+$ *prpF*$^+$)	100	27.0 ± 1.4
JE5584	JE3056/pBAD30	500	NG
JE7342	JE3056/pPRP151 (*S. oneidensis prpB*$^+$ *prpC*$^+$)	500	NG
JE5297	JE3946/pPRP21 (*S. oneidensis prpB*$^+$)	500	9.7 ± 0.5
JE7252	JE3946/pPRP149 (*S. oneidensis prpB*$^+$)	100	11.3 ± 0.1
JE7253	JE3946/pPRP151 (*S. oneidensis prpB*$^+$ *prpC*$^+$)	100	10.5 ± 0.5
JE5296	JE3946/pBAD30	500	NG
JE7270	JE3907/pPRP35 (*S. enterica prpC*$^+$)	100	11.0 ± 0.7
JE7254	JE3907/pPRP150 (*S. oneidensis prpC*$^+$)	500	12.6 ± 0.3
JE7255	JE3907/pPRP151 (*S. oneidensis prpB*$^+$ *prpC*$^+$)	500	11.6 ± 0.4
JE7251	JE3907/pBAD30	500	NG

[a] All plasmids except pPRP12-5.4 (*S. enterica prpBCD*$^+$) contain arabinose-inducible promoters. Where two plasmids are indicated, they are of different incompatibility groups. The following strains were used to construct the strains tested: JE3909 [*metE205 ara-9 prpD160 zai-6386*::Tn*10*d(Tc)], JE3907 [*metE205 ara-9 prpC107 zai-6386*::Tn*10*d(Tc)], JE3056 [*metE205 ara-9 prpB121*::Tn*10*d(Tc)], and JE3946 [*metE205 ara-9 prpB105 zai-6386*::Tn*10*d(Tc)].

[b] Each construct was tested in media with 100 and 500 µM arabinose; the highest growth rate was included in this table.

[c] JE7230 (JE3909/pPRP138 [*S. oneidensis acnD*$^+$]), JE7271 (JE3909/pPRP153 [*S. oneidensis prpF*$^+$]), JE7233 (JE3909/pPRP121 [*V. cholerae acnD*$^+$]), JE7273 (JE3909/pPRP155 [*V. cholerae prpF*$^+$]), JE7274 (JE3909/pPRP166 [*E. coli ybhHI*$^+$]), JE7356 (JE3909/pPRP163 [*E. coli ybhJ*$^+$]), JE7272 (JE3909/pPRP154 [*E. coli ybhH*$^+$]), or JE7359 (JE3909/pPRP154 [*E. coli ybhH*$^+$], pPRP163 [*E. coli ybhJ*$^+$]) did not grow in a propionate medium supplemented with 500 µM arabinose. NG, no growth observed (doubling times greater than 50 h).

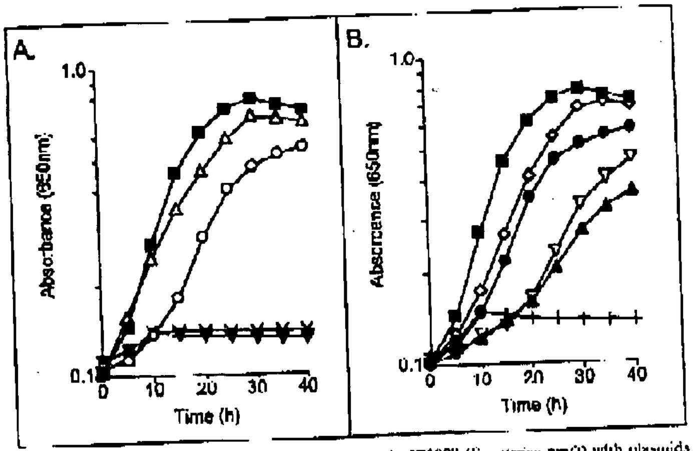


FIG. 3. Heterologous complementation studies. All curves shown are strain JE3909 (*S. enterica prpD*) with plasmids in *trans*. Filled squares, pPRP21 (*S. enterica prpD*$^+$); open triangles, pPRP123 (*V. cholerae acnD*$^+$ *prpF*$^+$); open circles, pPRP140 (*S. oneidensis acnD*$^+$ *prpF*$^+$); filled inverted triangles, pPRP138 (*S. oneidensis acnD*$^+$); X, pPRP153 (*S. oneidensis prpF*$^+$); open diamonds, pPRP121 (*V. cholerae acnD*$^+$) + pPRP153 (*S. oneidensis prpF*$^+$); filled circles, pPRP138 (*S. oneidensis acnD*$^+$) + pPRP153 (*S. oneidensis prpF*$^+$); open inverted triangles, pPRP130 (*S. oneidensis acnD*$^+$) + pPRP155 (*V. cholerae prpF*$^+$); filled triangles, pPRP121 (*V. cholerae acnD*$^+$) + pPRP155 (*V. cholerae prpF*$^+$); plus signs, pPRP163 (*E. coli ybhJ*$^+$) + pPRP154 (*E. coli ybhH*$^+$).

oneidensis PrpF. The *E. coli ybhJ* and *ybhH* genes were cloned individually and as an operon (*ybhHIJ*) under the control of an arabinose-inducible promoter. None of the plasmids tested restored the growth of strain JE3909 on propionate (Table 3). Strain JE3909 failed to grow on propionate when *E. coli ybhJ* was introduced with *S. oneidensis* or *V. cholerae prpF* or when *E. coli ybhH* was added with *S. oneidensis* or *V. cholerae acnD*. These data indicated that the *ybhHIJ* operon of *E. coli* did not convert the 2-MC generated by *S. enterica* into 2-MCA or at least did not convert enough to complement the growth phenotype on propionate. The role of the *ybhHIJ* operon in *E. coli* remains unclear.

***S. oneidensis* PrpB (2-MIC lyase) and PrpC (2-MC synthase) restore growth of *S. enterica prpB* and *prpC* mutant strains on propionate.** It was of interest to determine whether *S. oneidensis prpB* (a 2-MIC lyase ortholog) and *prpC* (a 2-MC synthase ortholog) could complement *S. enterica prpB* or *prpC* mutant strains. *S. oneidensis prpB* and *prpC* were cloned as a pair into plasmid pBAD18Kan and introduced into strains JE3946 (*prpB*) and JE3907 (*prpC*) (Table 3). The resulting *S. enterica* strains (JE7253 and JE7255, respectively) grew on propionate, suggesting that the same stereoisomer of 2-MC was synthesized in *S. enterica* and *S. oneidensis* and that the PrpB enzymes from both organisms most likely use the same stereoisomer of 2-MIC. It has been shown that *E. coli* only synthesizes the *2S,3S* isomer of 2-MC and generates the *2R,3S* isomer of 2-MIC (9), and because *S. enterica* PrpB and PrpC are greater than 91 and 96% identical, respectively, to the corresponding proteins in *E. coli*, it is inferred that the same stereoisomers of the 2-MC cycle intermediates are produced in these organisms. The same results were obtained when *S. oneidensis prpB* or *prpC* were cloned individually and introduced into *S. enterica* JE3946 (JE7252) and JE3907 (JE7254), respectively (Table 3).

AcnD and PrpF do not substitute for TCA cycle aconitase. It was also tested whether AcnD could substitute for *S. enterica* AcnA or AcnB during growth conditions that demanded a functional tricarboxylic acid (TCA) cycle. *S. oneidensis* or *V. cholerae acnD* individually or in combination with *prpF* were introduced into strain JE5993 (*acnA acnB*) (17). The resulting strains (JE6504, JE6506, JE7235, and JE7236) failed to grow on pyruvate and various concentrations of arabinose (as inducer). Only the control strains JE6502 (JE5993/pACN9 *S. enterica acnA*$^+$) and JE6503 (JE5993/pACN10 *S. enterica acnB*$^+$) grew on pyruvate under the conditions tested. Hence, it was concluded that the *acnD* and *prpF* genes cannot compensate for the lack of aconitase activity required for a functional TCA cycle of *S. enterica*, at least not to the level required for growth on pyruvate. No growth on pyruvate was observed when the *E. coli ybhJ* and *ybhH* genes or the complete *ybhHIJ* operon were introduced into strain JE5993.

Purification of AcnD and PrpF proteins. The putative biochemical activity of the AcnD and PrpF proteins was investigated in vitro. For this purpose, the *S. oneidensis acnD* and *prpF* genes were cloned and their products produced with an N-terminal chitin-binding tag (plasmids pPRP152 and pPRP156, respectively). Both proteins were purified by chitin affinity chromatography, and the tag was cleaved at the intein site according to the manufacturer's recommendations (New England Biolabs). Each protein was >95% pure as judged by scanning densitometry (Fig. 4). The relative mobility of both proteins was consistent with their predicted molecular masses, i.e., AcnD was observed at ~94 kDa, and PrpF was observed at ~42 kDa.

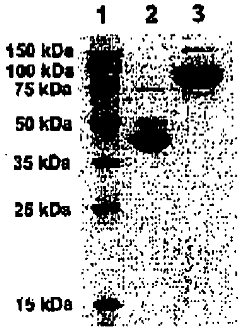


FIG. 4. SDS-PAGE of purified *S. oneidensis* AcnD and PrpF proteins. Lane 1, MW standards (Novagen); lane 2, purified PrpF; lane 3, purified AcnD.

Enzymatic activity of AcnD. Purified AcnD and H_6AcnA (17) were reactivated as described in Materials and Methods and assayed spectrophotometrically for activity with various substrates (Table 4). AcnD used citrate, *cis* aconitate, and 2-MC as substrates but not 2-MIC or isocitrate, indicating that, like PrpD (17), AcnD only catalyzes the first half of the aconitase-like reaction. The specific activity of AcnD for 2-MC was approximately 2.5-fold higher than that for citrate. The specific activity measured with 2-MC as substrate may be an underestimate of AcnD activity since commercially available 2-MC contained a mixture of stereoisomers, some of which may be inhibitory to the enzyme. On the other hand, AcnA used as positive control readily dehydrated 2-MIC, isocitrate, and citrate and hydrated *cis* aconitate. In agreement with previous work, AcnA did not use 2-MC as substrate (17).

AcnD is an Fe/S 2-MC dehydratase. AcnD activity was only observed after anoxic reactivation with iron, sulfide, and reductant. When reactivation was attempted in the absence of iron, no enzymatic activity was observed. The activity of reconstituted AcnD was lost over time in the presence of air. Also, when ferricyanide and EDTA were added to the enzyme after reactivation, all detectable activity was lost within 15 min of incubation. The primary amino acid sequence of AcnD contains 22 of the 23 residues found at the active site of mitochon-

TABLE 4. Specific activities of AcnD and AcnA[a] with different substrates

Substrate	AcnD	AcnA
Citrate	2.9 ± 0.3	26.3 ± 4.0
cis-Aconitate	5.0 ± 0.2	68.5 ± 7.4
Isocitrate	ND[b]	43.4 ± 3.1
2-MC	7.8 ± 0.4	ND
2-MIC	ND	12.7 ± 1.4

[a] AcnD and AcnA were reconstituted as stated in Materials and Methods.

[b] ND, no detectable activity observed (limits of detection were 0.5 μM/min for *cis*-aconitate conversion and 0.4 μM/min for 2-MCA formation) under the conditions tested.

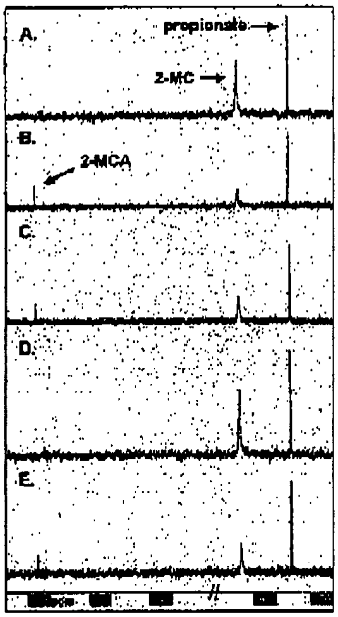


FIG. 5. ^{13}C-NMR spectra of in vitro reactions. The composition of the reaction mixtures is described under Materials and Methods. Reaction A contained the *S. enterica* PrpE and PrpC enzymes; reaction B contained the *S. enterica* PrpE, PrpC, and PrpD enzymes; reaction C contained the *S. enterica* PrpE, PrpC, and reactivated *S. oneidensis* AcnD enzymes; reaction D contained the *S. enterica* PrpE and PrpC enzymes and the *S. oneidensis* PrpF protein; reaction E contained the *S. enterica* PrpE and PrpC enzymes and the *S. oneidensis* AcnD enzyme and PrpF protein. Chemical shifts expressed in parts per million are as follows: [2-^{13}C]propionate 30.8 ppm; [2-^{13}C]MC, 47.5 ppm; and [2-^{13}C]MCA, 141.5 ppm. The portion of the spectra removed contained two glycerol peaks as previously described (17).

drial aconitase (mAcn), including the three cysteine residues that coordinate the 4Fe/4S cluster of mAcn (24). Taken together, these data indicated that in its active form AcnD contains an Fe/S center.

^{13}C-NMR analysis of the AcnD reaction product. [2-^{13}C]MC was synthesized in vitro as described (17). Reactivated AcnD was added to this reaction, incubated for 1 h, and analyzed by ^{13}C-NMR spectroscopy. Peak assignments were based on those identified previously (17). Excluding two glycerol peaks (glycerol was added to the enzyme as cryoprotectant), only three peaks were observed: for [2-^{13}C]propionate (30.8 ppm), [2-^{13}C]MC (47.5 ppm), and [2-^{13}C]MCA (141.5 ppm) (Fig. 5C). A positive control experiment for [2-^{13}C]MCA production was performed with H$_6$PrpD (Fig. 5B). The peaks observed with the H$_6$PrpD-catalyzed reaction matched exactly the shifts of those seen with the AcnD reaction. The reaction mixture containing only [2-^{13}C]MC had peaks at 30.8 and 47.5 ppm (Fig. 5A). These data indicated that both AcnD and PrpD had 2-MC dehydratase activity. To test whether the PrpF protein had aconitase-like enzymatic activity, PrpF was added to the [2-^{13}C]MC reaction mixture and incubated for 1 h at 37°C. No peak shift or decrease in the 2-MC signal at 47.5 ppm was observed (Fig. 5D). PrpF was also added to the reaction mixture containing reactivated AcnD. No peak shift or decrease in the 2-MCA signal at 141.5 ppm was observed relative to the signal observed in the reaction mixture containing only AcnD (Fig. 5E). These results indicated that PrpF did not convert 2-MCA into 2-MIC or catalyze the conversion of 2-MC into 2-MCA under the assay conditions tested.

Probing for a role for the PrpF protein in propionate metabolism. The possibility that the PrpF protein could catalyze the dehydration of citrate, isocitrate, 2-MC, or 2-MIC or the hydration of *cis*-aconitate was investigated. Even though PrpF does not contain an apparent Fe/S cluster binding motif, anoxic reconstitution of an Fe/S center was attempted. The PrpF protein with or without anoxic iron and sulfide reactivation did not have any detectable amounts of dehydratase or hydratase activities. PrpF was also added in twofold molar excess to AcnD in the *cis*-aconitate hydratase assay. No increase in the AcnD-catalyzed rate or in the overall conversion of substrate was observed.

We also tested whether AcnD required PrpF to catalyze the second half of the aconitase-like reaction, i.e., the conversion of 2-MCA to 2-MIC. PrpF was added to AcnD under anoxic and oxic conditions and was tested in the 2-MIC dehydratase assay. No 2-MIC dehydratase activity was observed.

We also looked into the possibility that PrpF could stabilize AcnD activity in the presence of oxygen. A twofold molar excess PrpF protein was added to anoxic AcnD. The mixture was incubated for 5 min before the seal was removed. Assays were performed, along with the control experiment with a reaction mixture that lacked PrpF protein. No significant differences in AcnD activity were observed over 2 h, and approximately 75% of AcnD (*cis*-aconitate hydratase) activity was lost over this period of time (data not shown). Work is currently being conducted to elucidate the role of PrpF in the 2-MC cycle.

DISCUSSION

This study established, both in vitro and in vivo, the biochemical activity of a new enzyme involved in the 2-MC cycle of several prokaryotes. The genes *acnD* and *prpF* of *S. oneidensis* and *V. cholerae*, when concurrently expressed, compensate for the lack of the Fe/S-independent PrpD enzyme in *S. enterica prpD* mutant strains during growth on propionate. The AcnD protein from *S. oneidensis* was isolated and shown to have a new activity for an enzyme containing an Fe/S center. AcnD catalyzes the dehydration of 2-MC and citrate but does not catalyze the dehydration of 2-MIC or isocitrate (Table 4). ^{13}C-NMR spectroscopy of reactivated AcnD with [2-^{13}C]MC revealed that AcnD could utilize the 2-MC generated by *S. enterica* PrpC and that the product of the AcnD reaction matched that produced by *S. enterica* PrpD (2-MCA) (Fig. 5). To our knowledge, this is the first report of an Fe/S-dependent 2-MC dehydratase.

To date, the only reported work on an AcnD homolog was performed with the *acnM* gene from *R. eutropha* (*R. eutropha* AcnM shares 83% identity with *S. oneidensis* AcnD) expressed in a crude extract system in *E. coli*. *E. coli* crude extracts containing *R. eutropha* AcnM protein were found to have *cis*-aconitate hydratase activity, but the data were inconclusive as to whether AcnM could dehydrate 2-MC. It was concluded, however, that AcnM may catalyze the hydration of 2-MCA into 2-MIC (8). In contrast, the data reported in this paper indicate that AcnD (and by extrapolation, AcnM) most likely does not catalyze the hydration of 2-MCA to 2-MIC because the enzyme will not dehydrate 2-MIC, and aconitases are known to catalyze freely reversible reactions. Additional support for this conclusion comes from NMR experiments where no evidence was obtained to indicate that active AcnD protein can convert 2-MC into 2-MIC. The only signal observed in the experiments was that of 2-MCA.

Studies on mammalian aconitases (mAcn) and AcnA and AcnB from both *S. enterica* and *E. coli* have demonstrated that these enzymes will not catalyze the dehydration of 2-MC; however, they will dehydrate 2-MIC and catalyze the full conversion of citrate into *cis*-aconitate into isocitrate (4, 9, 17, 29). The mechanism of aconitases is known to proceed by the binding of *cis*-aconitate in two ways to achieve the *trans* elimination or addition of water across the double bond (29). For this to occur, the substrate (*cis*-aconitate) must rotate 180°. The crystal structure of mAcn bound with 2-MIC allowed the prediction that if 2-MCA were rotated into the analogous 2-MC position, 2-MC would not be able to bind in the active site due to a steric clash of the methyl group with residue Asp165 (25). Interestingly, when aligned with mAcn, AcnD also contains this conserved aspartate residue, along with 21 of the other 22 active site residues of mAcn (24). It has been noted that all sequenced *acnD* homologs contain an Asn residue directly following one of the Cys residues that is likely to coordinate the Fe/S cluster, while in aconitases of the tricarboxylic acid cycle an Ile residue is found at this position (7). Experiments to test whether this residue plays a role in the substrate specificity of the AcnD enzyme are in progress.

It has been reported that horse heart aconitase can catalyze the formation of 2-MC and 2-MIC from 2-MCA (12). These data are in contradiction to the present information on aconitases, but the possibility exists that one enzyme may catalyze the complete conversion of 2-MC to 2-MIC. In *Y. lipolytica*, two enzymes, a 2-MC dehydratase and a 2-MIC dehydratase, have been implicated in its 2-MC cycle; both of these enzymes were found not to contain an Fe/S cluster or to be inactivated by iron-chelating agents (31, 34). To our knowledge, an Fe/S-independent 2-MIC dehydratase has not been identified in any other organism.

Possible roles for PrpF in propionate catabolism. The possibility that PrpF may be an isomerase of one of the intermediates of the 2-MC cycle was considered. However, it is unlikely that PrpF is a 2-MIC isomerase because the *prpB* gene from *S. oneidensis* complemented a *prpB* mutant strain of *S. enterica*, and NMR and spectrophotometric data suggested that AcnD only catalyzed the conversion of (2*S*,3*S*)-MC into 2-MCA. It was also possible that PrpF could be a 2-MC isomerase. However, our data showed that the *S. oneidensis* $prpC^+$ allele complemented an *S. enterica prpC* mutant strain, suggesting that the PrpC protein from *S. oneidensis* generated the same stereoisomer of 2-MC as that generated by the *S. enterica* PrpC enzyme. If the *S. oneidensis* PrpC protein generated a different 2-MC stereoisomer, it would be unlikely that the *S. enterica* PrpD enzyme would be able to use it as substrate. In support of this hypothesis, Brock et al. showed that the true substrate of the *E. coli* PrpD protein is most likely (2*S*,3*S*)-MC, and a 10-fold decrease in PrpD activity was observed when a racemic mixture of 2-MC stereoisomers was used (9). To further test these ideas, the stereochemical configuration of the reaction product of *S. oneidensis* PrpC must be determined.

Brämer et al. observed that PrpF was weakly similar (24%) to the *pduG* gene product of *S. enterica*, the proposed reactivation factor of diol dehydratase (7). Although this similarity is very weak, PrpF may be involved in AcnD Fe/S cluster formation or repair. As shown above, in vivo, PrpF must accompany AcnD to compensate for the lack of the PrpD enzyme during growth of a *prpD* mutant strain of *S. enterica* on propionate. If PrpF is required to stabilize AcnD in the presence of air, then growth under anoxic conditions should bypass the need for PrpF. However, PrpF was required even under anoxic growth conditions, suggesting that PrpF may not be needed to protect the AcnD Fe/S cluster from oxidation, but it could be involved in the formation, insertion, or stabilization of the Fe/S cluster. At present, the role of PrpF in propionate metabolism remains to be determined.

Bacteria use different strategies to convert 2-MC into 2-MIC. It is interesting that the sequenced *prp* operons of several bacteria contain both *prpD* and the *acnD*/*prpF* genes. Since these gene products catalyze the same reaction, i.e., the conversion of 2-MC into 2-MCA, they would appear to perform redundant functions in these organisms. Why would these organisms employ this strategy? Perhaps PrpD, the Fe/S-independent 2-MC dehydratase, is needed by these organisms at times when oxygen levels are high in the cell and may be deleterious to AcnD. Then why do these organisms retain the *acnD*/*prpF* pair of genes? Maybe having both of these ways to convert 2-MC into 2-MCA allows for more efficient growth on carbon sources that require the 2-MC cycle as a route of metabolism. Or perhaps the *acnD*/*prpF* gene products do carry out the second half of the aconitase-like reaction and convert 2-MCA into 2-MIC, which may not have been apparent in vitro in this study, in addition to the AcnD 2-MC dehydratase activity observed. If these gene products only function to convert 2-MC into 2-MCA, then an enzyme outside of the *prp* operons, most likely one of the aconitases of the cell (most bacteria contain more than one aconitase), must catalyze the conversion of 2-MCA into 2-MIC. It has been described for *Y. lipolytica* that an Fe/S-independent enzyme catalyzes the conversion of 2-MCA into 2-MIC (34). Perhaps some bacteria that utilize the 2-MC cycle also contain a gene encoding an Fe/S-independent 2-MIC dehydratase. Elucidating the function of PrpF both in vivo and in vitro may provide the answers to some of these questions.

ACKNOWLEDGMENTS

This work was supported by grant GM62203 from the NIGMS to J.C.E.-S. NMR spectroscopy studies were performed at the National Magnetic Resonance Facility at Madison, which is supported by the NIH Biomedical Technology Program (RR02301), with additional equipment funding from the University of Wisconsin, NSF Academic Infrastructure Program (BIR-9214394), NIH Shared Instrumentation Program (RR02781, RR08438), NSF Biological Instrumentation Program (DMB-8415048), and U.S. Department of Agriculture.

We thank A. R. Horswill for the AcnA protein, V. J. Starai for the PrpE protein, and W. W. Cleland for 2-MIC. We thank D. Saffarini and R. Taylor for their gifts of DNA.

REFERENCES

1. **Altschul, S. F., W. Gish, W. Miller, and E. W. Myers.** 1990. Basic local alignment search tool. J. Mol. Biol. 215:403–410.
2. **Aoki, H., H. Uchiyama, H. Umetsu, and T. Tabuchi.** 1995. Isolation of 2-methylcitrate dehydratase, a new enzyme serving in the methylcitric acid cycle for propionate metabolism, from *Yarrowia lipolytica*. Biosci. Biotechnol. Biochem. 59:1825–1828.
3. **Balch, W. E., and R. S. Wolfe.** 1976. New approach to the cultivation of methanogenic bacteria: 2-mercaptoethanesulfonic acid (HS-CoM)-dependent growth of *Methanobacterium ruminantium* in a pressurized atmosphere. Appl. Environ. Microbiol. 32:781–791.
4. **Beach, R. L., T. Aogaichi, and G. W. E. Plaut.** 1977. Identification of D-threo-α-methylisocitrate as stereochemically specific substrate for bovine heart aconitase and inhibitor of TPN-linked isocitrate dehydrogenase. J. Biol. Chem. 252:2702–2709.
5. **Berkowitz, D., J. M. Hushon, H. J. Whitfield, J. Roth, and B. N. Ames.** 1968. Procedure for identifying nonsense mutations. J. Bacteriol. 96:215–220.
6. **Bradford, M. M.** 1976. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Anal. Biochem. 72:248–255.
7. **Brämer, C. O., L. F. Silva, J. G. Gomez, H. Priefert, and A. Steinbüchel.** 2002. Identification of the 2-methylcitrate pathway involved in the catabolism of propionate in the polyhydroxyalkanoate-producing strain *Burkholderia sacchari* IPT101(T) and analysis of a mutant accumulating a copolyester with higher 3-hydroxyvalerate content. Appl. Environ. Microbiol. 68:271–279.
8. **Brämer, C. O., and A. Steinbüchel.** 2001. The methylcitric acid pathway in *Ralstonia eutropha*: new genes identified involved in propionate metabolism. Microbiology 147:2203–2214.
9. **Brock, M., C. Maerker, A. Schutz, U. Völker, and W. Buckel.** 2002. Oxidation of propionate to pyruvate in *Escherichia coli*: involvement of methylcitrate dehydratase and aconitase. Eur. J. Biochem. 269:6184–6194.
10. **Claes, W. A., A. Pühler, and J. Kalinowski.** 2002. Identification of two *prpDBC* gene clusters in *Corynebacterium glutamicum* and their involvement in propionate degradation via the 2-methylcitrate cycle. J. Bacteriol. 184:2728–2739.
11. **Davis, R. W., D. Botstein, and J. R. Roth.** 1980. A manual for genetic engineering: advanced bacterial genetics. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
12. **Gawron, O., and K. P. Mahajan.** 1966. Alpha-methyl-*cis*-aconitic acid. An aconitase substrate. II. Substrate properties and aconitase mechanism. Biochemistry 5:2343–2350.
13. **Grimek, T. L., H. M. Holden, I. Rayment, and J. C. Escalante-Semerena.** 2003. Residues C123 and D58 of the 2-methylisocitrate lyase (PrpB) enzyme of *Salmonella enterica* are essential for catalysis. J. Bacteriol. 185:4837–4843.
14. **Grimm, C., A. Evers, M. Brock, C. Maerker, G. Klebe, W. Buckel, and K. Reuter.** 2003. Crystal structure of 2-methylisocitrate lyase (PrpB) from *Escherichia coli* and modelling of its ligand bound active centre. J. Mol. Biol. 328:609–621.
15. **Gunsalus, R. P., S. M. Tandon, and R. S. Wolfe.** 1980. A procedure for anaerobic column chromatography employing an anaerobic Freter-type chamber. Anal. Biochem. 101:327–331.
16. **Guzman, L.-M., D. Belin, M. J. Carson, and J. Beckwith.** 1995. Tight regulation, modulation, and high-level expression by vectors containing arabinose P_{BAD} promoter. J. Bacteriol. 177:4121–4130.
17. **Horswill, A. R., and J. C. Escalante-Semerena.** 2001. In vitro conversion of propionate to pyruvate by *Salmonella enterica* enzymes: 2-methylcitrate dehydratase (PrpD) and aconitase enzymes catalyze the conversion of 2-methylcitrate to 2-methylisocitrate. Biochemistry 40:4703–4713.
18. **Horswill, A. R., and J. C. Escalante-Semerena.** 1997. Propionate catabolism in *Salmonella typhimurium* LT2: two divergently transcribed units comprise the *prp* locus at 8.5 centisomes, *prpR* encodes a member of the sigma-54 family of activators, and the *prpBCDE* genes constitute an operon. J. Bacteriol. 179:928–940.
19. **Horswill, A. R., and J. C. Escalante-Semerena.** 1999. The *prpE* gene of *Salmonella typhimurium* LT2 encodes propionyl-CoA synthetase. Microbiology 145:1381–1388.
20. **Horswill, A. R., and J. C. Escalante-Semerena.** 1999. *Salmonella typhimurium* LT2 catabolizes propionate via the 2-methylcitric acid cycle. J. Bacteriol. 181:5615–5623.
21. **Kennedy, M. C., M. H. Emptage, J. L. Dreyer, and H. Beinert.** 1983. The role of iron in the activation-inactivation of aconitase. J. Biol. Chem. 258:11098–11105.
22. **Kennedy, M. C., and H. Beinert.** 1988. The state of cluster SH and S^{2-} of aconitase during cluster interconversions and removal: a convenient preparation of apoenzyme. J. Biol. Chem. 263:8194–8198.
23. **Laemmli, U. K.** 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature 227:680–685.
24. **Lauble, H., M. C. Kennedy, H. Beinert, and C. D. Stout.** 1992. Crystal structures of aconitase with isocitrate and nitroisocitrate bound. Biochemistry 31:2735–2748.
25. **Lauble, H., and C. D. Stout.** 1995. Steric and conformational features of the aconitase mechanism. Proteins 22:1–11.
26. **Miyakoshi, S., H. Uchiyama, T. Someya, T. Satoh, and T. Tabuchi.** 1987. Distribution of the methylcitric acid cycle and β-oxidation for propionate catabolism in fungi. Agric. Biol. Chem. 51:2381–2387.
27. **Rado, T. A., and J. A. Hoch.** 1973. Phosphotransacetylase from *Bacillus subtilis*: purification and physiological studies. Biochim. Biophys. Acta 321:114–125.
28. **Sasse, J.** 1991. Detection of proteins, p. 10.6.1–10.6.8. *In* F. A. Ausubel, R. Brent, R. E. Kingston, D. D. Moore, J. G. Seidman, J. A. Smith, and K. Struhl (ed.), Current protocols in molecular biology, vol. 1. Wiley Interscience, New York, N.Y.
29. **Schloss, J. V., M. H. Emptage, and W. W. Cleland.** 1984. pH profiles and isotope effects for aconitases from *Saccharomycopsis lipolytica*, beef heart, and beef liver: α-methyl-*cis*-aconitate and *threo*-D_s-α-methylisocitrate as substrates. Biochemistry 23:4572–4580.
30. **Seidman, C. E., K. Struhl, J. Sheen, and T. Jessen.** 1997. Introduction of plasmid DNA into cells, p. 1.8.1–1.8.10. *In* F. M. Ausubel, R. Brent, R. E. Kingston, D. D. Moore, J. G. Seidman, J. A. Smith, and K. Struhl (ed.), Current protocols in molecular biology, vol. 1. Wiley Interscience, New York, N.Y.
31. **Tabuchi, T., H. Aoki, H. Uchiyama, and T. Nakahara.** 1981. 2-Methylcitrate dehydratase, a new enzyme functioning at the methylcitric acid cycle of propionate metabolism. Agric. Biol. Chem. 45:2823–2829.
32. **Tabuchi, T., and S. Hara.** 1974. Production of 2-methylcitric acid from *n*-paraffins by mutants of *Candida lipolytica*. Agr. Biol. Chem. 38:1105–1106.
33. **Tabuchi, T., and N. Serizawa.** 1975. The production of 2-methylisocitric acid from odd carbon *n*-alkanes by a mutant of *Candida lipolytica*. Agric. Biol. Chem. 39:1049–1054.
34. **Tabuchi, T., H. Umetsu, H. Aoki, and H. Uchiyama.** 1995. Characteristics of 2-methylisocitrate dehydratase, isolated from *Yarrowia lipolytica*, in comparison with aconitase. Biosci. Biotechnol. Biochem. 59:2013–2017.
35. **Textor, S., V. F. Wendisch, A. A. De Graaf, U. Müller, M. I. Linder, D. Linder, and W. Buckel.** 1997. Propionate oxidation in *Escherichia coli*: evidence for operation of a methylcitrate cycle in bacteria. Arch. Microbiol. 168:428–436.
36. **Thompson, J. D., D. G. Higgins, and T. J. Gibson.** 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. 22:4673–4680.
37. **Way, J. C., M. A. Davis, D. Morisato, D. E. Roberts, and N. Kleckner.** 1984. New Tn*10* derivatives for transposon mutagenesis and for construction of *lacZ* operon fusions by transposition. Gene 32:369–379.

JOURNAL OF BACTERIOLOGY, Dec. 2003, p. 7193–7201
0021-9193/03/$08.00+0 DOI: 10.1128/JB.185.24.7193–7201.2003
Vol. 185, No. 24
Copyright © 2003, American Society for Microbiology. All Rights Reserved.

A New Pathway for Salvaging the Coenzyme B_{12} Precursor Cobinamide in Archaea Requires Cobinamide-Phosphate Synthase (CbiB) Enzyme Activity

Jesse D. Woodson, Carmen L. Zayas, and Jorge C. Escalante-Semerena*

Department of Bacteriology, University of Wisconsin-Madison, Madison, Wisconsin

Received 5 June 2003/Accepted 8 September 2003

The ability of archaea to salvage cobinamide has been under question because archaeal genomes lack orthologs to the bacterial nucleoside triphosphate:5′-deoxycobinamide kinase enzyme (*cobU* in *Salmonella enterica*). The latter activity is required for cobinamide salvaging in bacteria. This paper reports evidence that archaea salvage cobinamide from the environment by using a pathway different from the one used by bacteria. These studies demanded the functional characterization of two genes whose putative function had been annotated based solely on their homology to the bacterial genes encoding adenosylcobyric acid and adenosylcobinamide-phosphate synthases (*cbiP* and *cbiB*, respectively) of *S. enterica*. A *cbiP* mutant strain of the archaeon *Halobacterium* sp. strain NRC-1 was auxotrophic for adenosylcobyric acid, a known intermediate of the de novo cobamide biosynthesis pathway, but efficiently salvaged cobinamide from the environment, suggesting the existence of a salvaging pathway in this archaeon. A *cbiB* mutant strain of *Halobacterium* was auxotrophic for adenosylcobinamide-GDP, a known de novo intermediate, and did not salvage cobinamide. The results of the nutritional analyses of the *cbiP* and *cbiB* mutants suggested that the entry point for cobinamide salvaging is adenosylcobyric acid. The data are consistent with a salvaging pathway for cobinamide in which an amidohydrolase enzyme cleaves off the aminopropanol moiety of adenosylcobinamide to yield adenosylcobyric acid, which is converted by the adenosylcobinamide-phosphate synthase enzyme to adenosylcobinamide-phosphate, a known intermediate of the de novo biosynthetic pathway. The existence of an adenosylcobinamide amidohydrolase enzyme would explain the lack of an adenosylcobinamide kinase in archaea.

To date, de novo coenzyme B_{12} (Fig. 1) biosynthesis has only been reported to occur in prokaryotes (2, 13, 28, 30, 31, 38). This major biosynthetic pathway has mostly been studied in bacterial systems, with the majority of the work being focused on the anaerobic biosynthesis of the corrin ring in *Salmonella enterica* (11, 27), *Propionibacterium freudenreichii* subsp. *shermanii* (29), and *Bacillus megaterium* (6, 23, 24) and on aerobic biosynthesis of the corrin ring in *Pseudomonas denitrificans* (4). This large body of work has given considerable insight into the details of cobamide biosynthesis and has set the basis for comparisons with other organisms (26, 38).

At present, our knowledge of how archaea synthesize cobamides is very limited (7, 36, 39). It is clear that some archaea synthesize and require cobamides to live. For example, methanogenic archaea require cobamides for methanogenesis from H_2 and CO_2, acetate, or methanol (10). The extremely halophilic archaeon *Halobacterium* sp. NRC-1 has been shown to produce and require cobamides under certain growth conditions, but it is unclear why they are needed (39). Some archaea may possess cobamide-dependent ribonucleotide reductases that are required for DNA synthesis, as suggested by genome sequence analysis. In fact, cobamide-dependent ribonucleotide reductases have been isolated from *Thermoplasma acidophilum* and *Pyrococcus furiosus* (25, 34). The availability of several archaeal genome sequences has allowed researchers to predict which organisms may have complete de novo cobamide pathways and which may have only enough genetic information for precursor salvaging.

Analysis of the available archaeal genome sequences revealed the absence of an archaeal ortholog to the bacterial ATP:adenosylcobinamide (AdoCbi) kinase/GTP:adenosylcobinamide-phosphate (AdoCbi-P) guanylyltransferase (CobU in *S. enterica*). The transferase activity was shown to be required for de novo biosynthesis of cobamides and for the salvaging of unphosphorylated Cbi (19). The kinase activity, on the other hand, is only required for the salvaging of Cbi (8, 36) (Fig. 1). Recently, it was shown that the conserved archaeal *cobY* gene is the nonorthologous replacement of the *S. enterica cobU* gene. The CobY protein has the nucleoside triphosphate (NTP):AdoCbi-P nucleotidyltransferase activity required for de novo synthesis of cobamides but lacks the NTP:AdoCbi kinase activity necessary to salvage Cbi via the pathway used by bacteria (5, 36, 39).

The lack of an NTP:AdoCbi kinase ortholog in archaea raises three important questions. (i) Are archaea able to salvage Cbi? (ii) If they can, does an alternative, nonorthologous replacement of the bacterial NTP:AdoCbi kinase exist in these prokaryotes? (iii) If a nonorthologous replacement of the bacterial NTP:AdoCbi kinase does not exist in archaea, does an alternative, uncharacterized Cbi-salvaging pathway exist? Previous studies of *Methanobacterium thermoautotrophicum* strongly suggested that this archaeon can salvage Cbi (32). However, to the best of our knowledge, there are no reported studies of the pathway used by this or any other archaeon to salvage Cbi.

---

* Corresponding author. Mailing address: Department of Bacteriology, University of Wisconsin, 264 Enzyme Institute, 1710 University Ave., Madison, WI 53726-4087. Phone: (608) 262-7379. Fax: (608) 265-7909. E-mail: escalante@bact.wisc.edu.

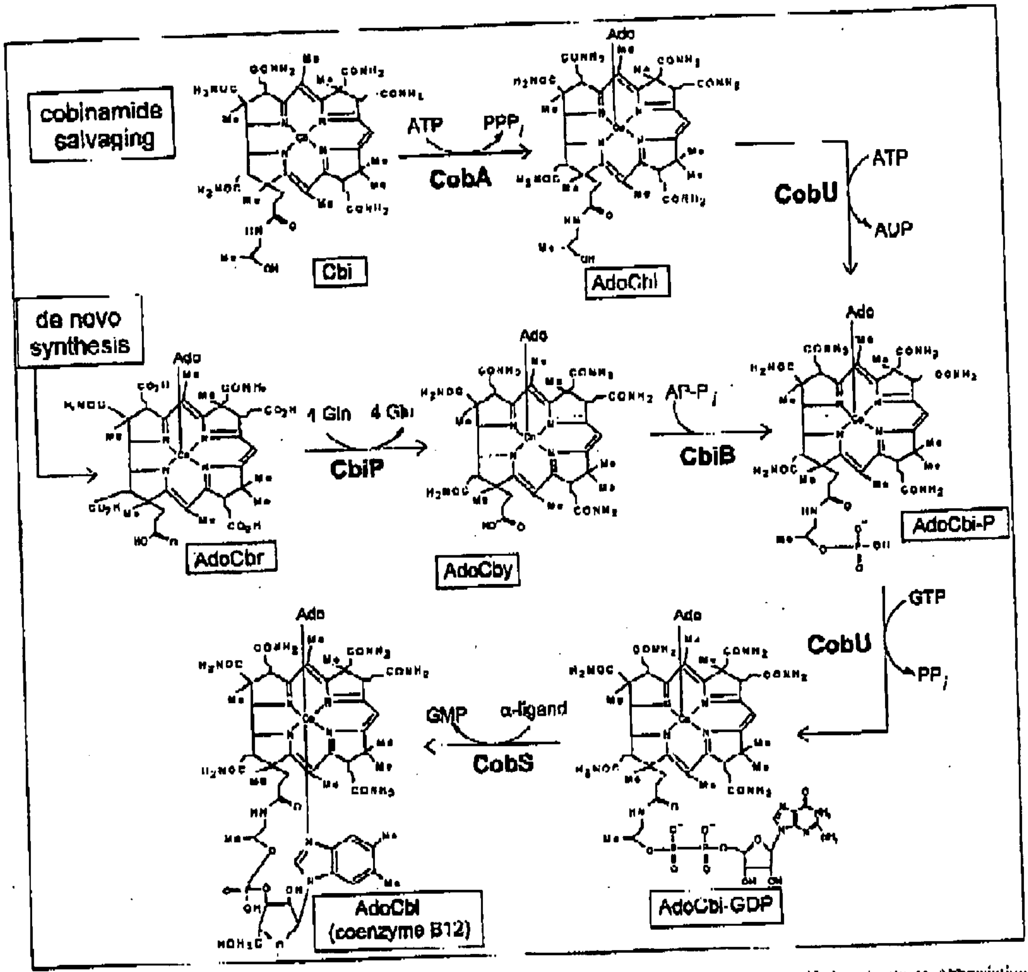


FIG. 1. Late steps of cobamide biosynthesis in the bacterium *S. enterica*. Intermediates are boxed and indicated below structures. Abbreviations: AP-P$_i$, aminopropanol phosphate; AdoCbr, adenosylcobyrinic acid *a,c*-diamide; AdoCbl, adenosylcobalamin, CobS, cobalamin (5′-*P*) synthase.

In this paper, we provide genetic evidence for the ability of the extremely halophilic archaeon *Halobacterium* sp. strain NRC-1 to efficiently salvage exogenous Cbi via an alternative pathway to the one used by bacteria. These studies demanded the functional characterization of two genes whose putative function had been annotated exclusively on the basis of their homology to the bacterial adenosylcobyric acid (AdoCby) and AdoCbi-P synthases (*cbiP* and *cbiB*, respectively) present in *S. enterica* (Fig. 1).

MATERIALS AND METHODS

Strains and plasmids. The genotypes of the *Halobacterium* sp. strain NRC-1 and *S. enterica* strains and the plasmids used in this work are described in Table 1.

Chemicals, culture media, and growth conditions. All chemicals used in this work were commercially available, high-purity compounds. When corrinoids were added to the medium, they were used at concentrations of 100 pM for *Halobacterium* studies and 15 nM for *S. enterica* studies. All corrinoids were added in their cyano form. Cbi dicyanide was purchased from Sigma (St. Louis, Mo.). Cbi-GDP dicyanide was synthesized as previously described (36). Cobyric acid dicyanide [$(CN)_2Cby$] was a gift from Paul Renz (Universität-Hohenheim, Stuttgart, Germany). 5-Fluoroorotic acid (5-FOA) was purchased from Zymo Research (Orange, Calif.), and mevinolin was purchased from LKT Laboratories, Inc. (St. Paul, Minn.).

***Halobacterium* studies.** Strains were grown in liquid peptone (Oxoid, Hampshire, England) medium (18) lacking trace metals. *Halobacterium* cultures were grown to stationary phase at 37°C with shaking for 5 days. Cells used as inocula were harvested by centrifugation (10,000 × *g* for 2 min) with a Microfuge 18 centrifuge (Beckman-Coulter, Fullerton, Calif.) and washed once in a chemically

TABLE 1. Strains and plasmids used in this study

Strain or plasmid	Marker(s)[a]	Relevant genotype	Description	Reference or source[b]
Halobacterium strains				
MPK414		Δ*ura3*	Strain with de novo cobamide biosynthetic capability	39
JE6738		Δ*ura3* Δ*cbiP*	Strain with in-frame deletion of *cbiP*	
JE6791		Δ*ura3* Δ*cbiB*	Strain with in-frame deletion of *cbiB*	
JE6930		Δ*ura3* Δ*cbiB* *ura3*::*cbiB*$^+$	Strain used to test for complementation of *cbiB*	
JE7001		Δ*ura3* Δ*cbiP* *ura3*::*cbiP*$^+$	Strain used to test for complementation of *cbiP*	
S. enterica strains				
TR6583		*metE*	*S. enterica* wild type for this study	Laboratory collection
JE588		*cbiP metE*	*S. enterica* strain used for *cbiP* complementation studies	Laboratory collection
JE6368		*cbiB metE*	*S. enterica* strain used for *cbiB* complementation studies	Laboratory collection
Plasmids				
pMPK428	5-FOA^s, Mev^r	*ura3*$^+$	Plasmid used to generate in-frame deletions of targeted genes	22
pMPK424	5-FOA^s, Mev^r	*ura3*$^+$	Plasmid contains flanking sequence to *ura3* to allow recombination at the *ura3* locus	21
pCBIP2	5-FOA^s, Mev^r	*ura3*$^+$ Δ*cbiP*	Plasmid transformed into MPK414 to delete *cbiP*	
pCBIP7	5-FOA^s, Mev^r	*ura3*$^+$ *cbiP*$^+$	Plasmid used to recombine *cbiP* into *ura3* locus	
pVng1578-2	5-FOA^s, Mev^r	*ura3*$^+$ Δ*cbiB*	Plasmid transformed into MPK414 to delete *cbiB*	
pVng1578-3	5-FOA^s, Mev^r	*ura3*$^+$ *cbiB*$^+$	Plasmid used to recombine *cbiB* into *ura3* locus	
pT7-7	Ap^r		Cloning vector used for complementation studies in *S. enterica*	33
pCBIP9	Ap^r	*cbiP*$^+$	Plasmid used to provide *S. enterica* *cbiP* in *trans*	
pMmCBIP1	Ap^r	*cbiP*$^+$	Plasmid used to provide *M. mazei* *cbiP* in *trans*	
pScCBIB4	Ap^r	*cbiB*$^+$	Plasmid used to provide *S. enterica* *cbiB* in *trans*	Laboratory collection
pMmCBIB1	Ap^r	*cbiB*$^+$	Plasmid used to provide *M. mazei* *cbiB* in *trans*	

[a] Abbreviations: Mev^r, resistance to mevinolin; 5-FOA^s, sensitivity to 5-fluoroorotic acid; Ap^r, resistance to ampicillin.
[b] Unless otherwise stated, strains and plasmids were constructed during the course of this study.

defined medium (14). Cells were diluted 100-fold and used to inoculate the defined medium containing the appropriate corrinoid supplements. Cultures were grown at 37°C with shaking. Growth was monitored every 24 h by measuring the absorbance of the culture at 650 nm with a Spectronic 20D spectrophotometer (Milton Roy, Rochester, N.Y). In all cases, media were supplemented with uracil (450 μM).

***S. enterica* studies.** Plasmids were introduced into *S. enterica* by passing them first through a restriction deficient strain (37).

Anaerobic growth studies. Four independent colonies of each strain were patched onto Luria-Bertani–ampicillin (100 μg/ml) agar (1.5%), grown for 5 h at 37°C, and replica printed onto defined, no-carbon E medium (3) supplemented with glucose (11 mM), $MgSO_4$ (1 mM), 1,2-propanediol (10 mM), CoCl (5 μM), ampicillin (25 μg/ml), and trace minerals (1). $(CN)_2$Cby was added as indicated. Plates were incubated anaerobically in an ANA PAK system (Scott Laboratories, Inc., Fiskeville, R.I.), with a BBL GasPak anaerobic system (Becton Dickinson, Cockeysville, Md). The growth of the strains after 24 h indicated de novo cobamide biosynthesis.

Aerobic growth studies. *S. enterica* strains were grown to full density in nutrient broth (Difco) supplemented with ampicillin (100 μg/ml). Cells were diluted 100-fold and used to inoculate the defined no-carbon E medium supplemented with glucose (11 mM), $MgSO_4$ (1 mM), 1,2-propanediol (10 mM), ampicillin (25 μg/ml), and trace minerals (1). Corrinoid supplements were added as indicated. Cultures were monitored while grown at 37°C with continuous shaking (14 Hz) in an EL808 Ultra Microplate Reader (Bio-Tek Instruments, Inc., Winooski, Vt)

Plasmid constructions. Plasmids were propagated in the *Escherichia coli* strain DH5α except where noted. In all cases, *Halobacterium* sp. strain NRC-1 genomic DNA for PCR was prepared as previously described (39). *Methanosarcina mazei* strain Gö1 DNA for PCR was a gift from Gerhard Gottschalk (Göttingen, Germany). All primers were purchased from Integrated DNA Technologies, Inc. (Coralville, Iowa). Underlined portions of the primer sequences (see below) indicate introduced restriction sites.

***Halobacterium* plasmids.** A diagram of the *Halobacterium* sp. strain NRC-1 DNA included in the most relevant plasmids is included in Fig. 2B.

(i) Plasmid pCBIP1. The 5′ primer cbiPΔ*Hin*dIII5′#2 (GTTCGGCAAAAGCTTCCCACCGCAG) and the 3′ reverse primer cbiPΔ*Eco*RV3′ (CTGGACTCGATATCGTTGAGCAAC) were used to amplify an 814-bp PCR fragment from strain MPK414 genomic DNA. Amplified DNA was cut with *Hin*dIII/*Eco*RV restriction enzymes (unless otherwise noted, the underlined portion of the sequence is the restriction enzyme site), purified with a QIAquick gel extraction kit (QIAGEN; Valencia, Calif.), and cloned into the *Hin*dIII/*Sma*I restriction site of plasmid pMPK428, which contains the wild type allele of the *Halobacterium* sp. *ura3* gene and a mevinolin resistance determinant (32). The resulting plasmid is referred to as pCBIP1.

(ii) Plasmid pCBIP2. Plasmid pCBIP2 (Δ*cbiP* *ura3*$^+$) carries an in-frame deletion of the *Halobacterium* sp. strain NRC-1 *cbiP* gene and was constructed as follows. The 5′ primer cbiPΔ*Xba*I5′ (GCACGTGCTCTAGATGATGAAAG) and reverse 3′ primer cbiPΔ*Hin*dIII3′ (CACGACGAGTAAGCTTTCGCCTTC) were used to amplify an 807 bp fragment from MPK414 genomic DNA.

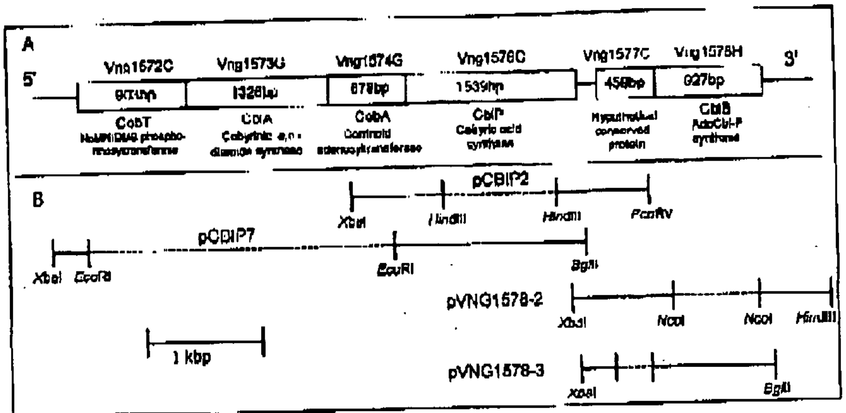


FIG. 2. Putative operons in *Halobacterium* sp. strain NRC-1 containing *cbiP* (Vng1576G) and *cbiB* (Vng1578H) and plasmid constructions. (A) The reported ORF designation is shown above each rectangle with our annotation below it. The reported length (base pairs) of each ORF is indicated within each box. (B) Brackets connected by solid lines indicate the regions of DNA that were included in plasmids pCBIP2, pCBIP7, pVNG1578-2, and pVNG1578-3. Dashed lines indicate regions that were not included in the plasmids. The DNA restriction enzyme sites used for cloning purposes are labeled below the brackets.

The fragment was cut with *XbaI/HindIII* restriction enzymes, gel purified, and cloned into the *XbaI/HindIII* restriction site of plasmid pCBIP1 to create plasmid pCBIP2. The latter contained an in-frame deletion of *cbiP* that replaced bases 303 to 1376 with a 6-bp *HindIII* restriction site, thus deleting 358 of the 512 amino acids. Plasmid pCBIP2 also carries the mevinolin resistance determinant and a wild-type allele of the *ura3* gene.

(iii) Plasmid pCBIP4. The 5' primer cbiPComp*EcoRI*5' (TCTAGAGAATT CTATGTCGACGTTCGTGACCGAG) and reverse primer cbiPComp*BglII*3' (A GATCTTAGATTTTAAAAGCCGCCCGGTTCAAACCACGTTGACACGG TAG) were used to amplify a 1,749-bp PCR product from strain MPK414 genomic DNA. The fragment was cloned into pGEM-T with the Promega pGEM-T cloning kit (Madison, Wis.) to yield the plasmid pCBIP4.

(iv) Plasmid pCBIP5. The fragment carried on plasmid pCBIP4 was excised as a 1,721-bp fragment with an *EcoRI/BglII* digest, gel purified, and cloned into the *EcoRI/BglII* restriction site of p17-7 (33) to yield plasmid pCBIP5.

(v) Plasmid pCBIP6. The 5' primer cbiPComp*Xba*I5' (TCTAGATCTAGAC CCAACTCTGGTTGCATACAG) and reverse primer cbiPComp*EcoRI*3' (GAA TTCGAATTCGCCCTACGTCAGCACTTCG) were used to amplify a 286-bp PCR product from strain MPK414 genomic DNA. The fragment was cut with *XbaI/EcoRI* restriction enzymes, gel purified, and cloned into the *XbaI/EcoRI* restriction site of plasmid pCBIP5 to yield plasmid pCBIP6.

(vi) Plasmid pCBIP7. The 268-bp *XbaI/EcoRI* and 1,721-bp *EcoRI/BglII* fragments from plasmid pCBIP6 were excised as a single 1,989-bp fragment with *XbaI/BglII* restriction enzymes, gel purified, and cloned into the *XbaI/BglII* restriction site of plasmid pMPK424 (21), which was prepared from the *dam* mutant strain GM2163 (New England Biolabs, Manchester, Mass.) to yield plasmid pCBIP7 (*ura3*$^+$ *cbiP*$^+$). Plasmid pCBIP7 contained the 1,989-bp fragment flanked by a sequence that would allow recombination at the *Halobacterium* sp. strain NRC-1 *ura3* locus. The resulting plasmid carried a wild-type copy of the *cbiP* gene, including 107 bases 5' of the putative start codon and 210 bases upstream of the putative operon. To include these sequences, parts of the Vng1572C and Vng1574G open reading frames (ORFs) were also cloned, but the segments carried an in-frame fusion that fused amino acid residue 15 (of 300) of Vng1572C to residue 191 (of 225) of Vng1574G with Glu and Phe encoded by the introduced *EcoRI* site (Fig. 2B). Including these sequences should preserve the regulation of *cbiP* in its own operon without including other genes. Flanking the 3' end was a 16-bp sequence derived from the *bop* transcription terminator sequence (9) to ensure termination of the *cbiP* mRNA transcript.

(vii) Plasmid pVNG1578-1. The 5' primer Vng1578*NcoI*5' (CCATGGCCAT GGCGTTGTTCTACGGCGGAGGTGG) and 3' reverse primer Vng1578*HindIII*3' (AAGCTTAAGCTTACTTCGAACAGCGGCTTCTCG) were used to amplify an 855-bp PCR fragment from strain MPK414 genomic DNA. The fragment was cut with *NcoI/HindIII* restriction enzymes, gel purified, and cloned into the *NcoI/HindIII* restriction site of plasmid pMPK428, which contains the wild-type allele of *Halobacterium* sp. strain NRC-1 *ura3* and a mevinolin resistance determinant (22). The resulting plasmid is referred to as pVng1578-1.

(viii) Plasmid pVNG1578-2. Plasmid pVNG1578-2 (Δ*cbiB ura3*$^+$) carried an in-frame deletion of the *Halobacterium* sp. strain NRC-1 *cbiB* gene and was constructed as follows. The 5' primer Vng1578*Xba*I5' (TCTAGATCTAGACG CCCACGTCGTATTCGACC) and reverse 3' primer Vng1578*NcoI*3' (CCAT GGCCATGGCCTCCACGGTTTCTTCGACG) were used to amplify an 841-bp fragment from MPK414 genomic DNA. The fragment was cut with *XbaI/NcoI* restriction enzymes, gel purified, and cloned into the *XbaI/NcoI* restriction site of plasmid pVNG1578-1 to create plasmid pVNG1578-2. The latter contained an in-frame deletion of *cbiB* that replaced bases 133 to 897 with a 6-bp *NcoI* restriction site, thus deleting 255 of the 308 amino acids. Plasmid pVNG1578-2 also carries the mevinolin resistance determinant and a wild-type allele of the *ura3* gene.

(ix) Plasmid pVNG1578-3. The plasmid pVng1578-3 (*cbiB*$^+$ *ura3*$^+$) carries a wild-type allele of the *Halobacterium* sp. strain NRC-1 *cbiB* gene and was constructed as follows. The 5' primer cbiBComp*Xba*I5' (GAATCCTCTAGATTGA CCGACCCGATTCAAGTCC) and the reverse primer cbiBComp*BglII*3' (GAAT TCAGATCTAAAACCCGCGCCGGTTGGCGATTGAACGCCTCCCAG) were used to amplify a 1,398-bp PCR product from strain JE6493 (a derivative of MPK414) with an in-frame deletion on Vng1577, deleting bases 103 to 408 (J. C. Escalante-Semerena, laboratory collection) genomic DNA. The fragment was cut with *XbaI/BglII* restriction enzymes, gel purified, and cloned into the *XbaI/BglII* restriction site of plasmid pMPK424 (21) (prepared from the mutant strain GM2163 *dam*) (New England Biolabs, Manchester, Mass.) to yield plasmid pVNG1578-3 (*ura3*$^+$ *cbiB*$^+$). The latter contains the cloned fragment flanked by a sequence that would allow recombination at the *ura3* locus of *Halobacterium* sp. strain NRC-1. The resulting plasmid carried a wild-type copy of the *cbiB* gene, including 47 bases upstream of the putative start codon and 200 bases upstream of the putative operon. To include these sequences, part of ORF Vng1577C was also cloned, but it carried an in-frame deletion spanning from residue 35 to residue 136 (of 152). Including these sequences should preserve the regulation of *cbiB* in its own operon without including other genes. Flanking the 3' end was a 16-bp sequence derived from the *bop* transcription terminator sequence (9) to ensure transcriptional termination of the *cbiB* mRNA transcript.

S. enterica **plasmid pCBIP9.** The plasmid pCBIP9 contained a wild-type allele of *S. enterica cbiP* under the control of the *lac* promoter and ribosome binding site and was constructed as follows. The fragment carried on plasmid pCBIP3 (Escalante-Semerena, laboratory collection) included only the *S. enterica cbiP* ORF and was excised as a 1,520-bp fragment with an *NdeI/XhoI* digest, gel purified, and cloned into the *NdeI/SalI* restriction site of p17-7 (33) to produce plasmid pCBIP9(*cbiP*$^+$).

***M. mazei* plasmids.** (i) Plasmid pMmCBIP1. Plasmid pMmCBIP1 (*cbiP*$^+$) contained a wild-type allele of *M. mazei* strain Goe1 *cbiP* (ORF Mma0093) under the control of the *lac* promoter and ribosome-binding site and was constructed as follows. The 5' primer Mma0093-Blunt#1 (TGAATAATAAAAACCCTGTTT GCCCAG) and the reverse primer Mma0093-*Sal*I-3' (CGCGTCGTCGACTC AGACTCCTCC) were used to amplify a 1,513 bp PCR product from *M. mazei* genomic DNA. The fragment was treated with polynucleotide kinase, cut with *Sal*I, gel purified, and cloned into the *Nde*I/*Sal*I site of pT7-7 (prepared by cutting plasmid pT7-7 with *Nde*I, blunt ending with the MBI Fermentas [Amherst, N.Y.] DNA polymerase I large (Klenow) fragment, and digesting with *Sal*I to produce the plasmid pMmCBIP1 (*cbiP*$^+$)).

(ii) Plasmid pMmCBIB1. Plasmid pMmCBIB1 (*cbiB*$^+$) contained a wild-type allele of *M. mazei* strain Goe1 *cbiB* (ORF Mma2050) under the control of the *lac* promoter and ribosome-binding site and was constructed as follows. The 5' primer MmcbiB-5'*Nde*I #2 (5'-AGCCTATCATATGATCATACCGGACAGC-3') and the reverse primer MmcbiB 3' *Sal*I (5'-ATTGATCTGGAGTAAGTCG ACTTTTCAGGG-3') were used to amplify a 1,025-bp PCR product from *M. mazei* genomic DNA. The fragment was cut with *Nde*I/*Sal*I restriction enzymes, gel purified, and cloned into the *Nde*I/*Sal*I restriction site of plasmid pT7-7 to produce plasmid pMmCBIB1 (*cbiB*$^+$).

***Halobacterium* strain constructions.** (i) Construction of a Δ*cbiP* mutant strain. An in-frame deletion of *cbiP* in the chromosome of strain MPK414 (Δ*ura3*) was generated by using previously described methodology (20). Briefly, strain JE6738 (Δ*ura3* Δ*cbiP*) was constructed by transforming strain MPK414 with plasmid pCBIP2 as described previously (15). Flanking sequences of over 200 bases on each side of the deleted *cbiP* gene ensured efficient recombination of the fragment into the chromosome. Mevinolin-resistant transformants were selected as described previously (15) and replated on medium containing 5-FOA to select for the loss of the plasmid (20). Colonies resistant to 5-FOA were screened by PCR to identify the desired recombinant (Δ*cbiP*). DNA sequencing was used to confirm the in-frame deletion of the *cbiP* gene in the chromosome of strain JE6738

(ii) Construction of a Δ*cbiB* mutant strain. An in-frame deletion of *cbiB* in the chromosome of strain MPK414 was generated by using the same strategy as mentioned above. Strain JE6791 (Δ*ura3* Δ*cbiB*) was constructed with strain MPK414 and plasmid pVNG1578-2. DNA sequencing was used to confirm the in-frame deletion of the *cbiB* gene in the chromosome of strain JE6791.

(iii) Construction of a *cbiP* complementation strain. Complementation studies were performed with a single copy of the wild-type allele of the gene in question placed at the *ura3* locus. For *cbiP* complementation studies, a wild-type allele of *cbiP* was placed at the chromosomal *ura3* locus of strain JE6738. Plasmid pCBIP7 was transformed into strain JE6738, and strains carrying the *cbiP*$^+$ allele at the chromosomal *ura3* locus (strain JE7001 [Δ*ura3* Δ*cbiP ura3*::*cbiP*$^+$]) were isolated by using the same *ura3*-based gene replacement method for the isolation of deleted genes. PCR and DNA sequencing verified the presence of *cbiP*$^+$ at the *ura3* locus.

(iv) Construction of a *cbiB* complementation strain. For *cbiB* complementation studies a wild-type allele of *cbiB* was placed at the chromosomal *ura3* locus of strain JE6701. Plasmid pVNG1578-3 was transformed into strain JE6791, and a strain carrying the *cbiB*$^+$ allele at the chromosomal *ura3* locus (strain JE6930 [Δ*ura3* Δ*cbiB ura3*::*cbiB*$^+$]) was isolated. PCR and DNA sequencing verified the presence of *cbiB*$^+$ at the *ura3* locus.

RESULTS

Rationale used to probe into corrinoid salvaging in *Halobacterium*. Because the growth of *Halobacterium* in defined medium requires cobamides, the growth of a corrinoid-deficient mutant in medium supplemented with incomplete cobamide precursors would be indicative of precursor salvaging. To block corrin ring biosynthesis in *Halobacterium*, in-frame deletions were introduced in the second-to-last step or in the last step of corrin ring biosynthesis. In *S. enterica*, these steps of the pathway are catalyzed by the AdoCby synthase (CbiP) enzyme and the AdoCbi-P synthase (CbiB) enzyme, respectively (38). It was hypothesized that a block in either one of these steps would render a strain dependent on exogenous Cby or Cbi precursors. The mutation in *cbiP* would block salvaging of cobyrinic acid *a,c*-diamide but should not interfere with Cby or Cbi salvaging. A mutation in *cbiB* would address the question of what the point of entry of Cbi is in the *Halobacterium* genome sequence. That is, if a *cbiB* mutation does not prevent Cbi salvaging, then an unidentified kinase may be responsible for the activation of Cbi to Cbi-P (the substrate of the CobY enzyme). Alternatively, the inability of a *cbiB* mutant to salvage Cbi would suggest the existence of a new pathway for the activation of Cbi in this archaeon.

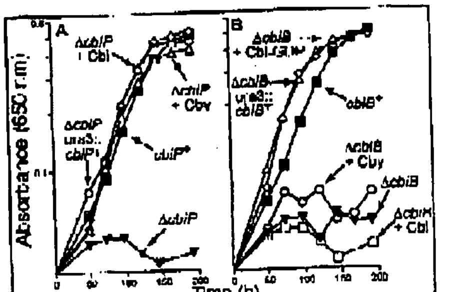


FIG. 3. Nutritional studies of *Halobacterium* sp. strain NRC-1 strains. Cobamide-dependent growth of *Halobacterium* sp. strain NRC-1 strains in the defined liquid medium at 37° is reported as absorbance at 650 nm as a function of time. The strains are indicated by their genotypes. The corrinoids added to the medium are indicated next to the genotypes. The strains used were MPK414 *cbiP*$^+$ *cbiB*$^+$, JE6738 Δ*cbiP*, JE7001 Δ*cbiP ura3*::*cbiP*$^+$, JE6791 Δ*cbiB*, and JE6930 Δ*cbiB ura3*::*cbiB*$^+$. Abbreviations: Cby, cobyric acid dicyanide; Cbi, cobinamide dicyanide; Cbi-GDP, cobinamide-GDP dicyanide. In all cases, corrinoids were used at concentrations of 100 pM.

Identification of the *cbiP* and *cbiB* genes of *Halobacterium*. ORF Vng1576G (gene identification [gi] number 15790548) of the *Halobacterium* sp. strain NRC-1 genome sequence (17) was identified as the putative *cbiP* gene of this archaeon based on the 40% identity and 53% similarity of the predicted gene product to the CbiP protein of *S. enterica*. In the *Halobacterium* genome, the *cbiP* (ORF Vng1576G) gene is located at the 3' end of a putative operon containing ORF Vng1574G and ORF Vng1573G, which encode the putative orthologs of the bacterial ATP:co(I)rrinoid adenosyltransferase (CobA in *S. enterica*) and the cobyrinic acid *a,c* diamide synthase (CbiA in *S. enterica*), respectively (Fig. 2A). These two proteins are believed to modify the corrinoid immediately preceding the CbiP-catalyzed step (38).

ORF Vng1578H (gi number 15790550) of the *Halobacterium* genome sequence was identified as the putative *cbiB* gene of this archaeon based on the 30% identity and 43% similarity of the predicted gene product to the CbiB of *S. enterica*. In the *Halobacterium* genome, the *cbiB* gene is the promoter-distal gene in a putative operon containing one other ORF of unknown function (Fig. 2A).

***cbiP* (ORF Vng1576G) is a cobamide biosynthetic gene in *Halobacterium*.** To determine if strain JE6738 (Δ*cbiP*) was deficient in cobamide biosynthesis, growth was assessed in defined medium where cobamides were essential for growth. Unlike strain MPK414 (*cbiP*$^+$), strain JE6738 (Δ*cbiP*) failed to grow in the defined medium lacking corrinoids (Fig. 3A). To determine if the observed lack of growth of JE6738 was caused by the inability to synthesize cobamides de novo, the medium was supplemented with Cby (the nonadenosylated product of the CbiP-catalyzed reaction). The addition of Cby restored wild-type growth of JE6738 (Fig. 3A) but did not significantly enhance the growth of the wild-type strain (data not shown). The doubling times of strains MPK414 and JE6738 in medium supplemented with Cby were very similar (30 and 27 h, respectively), whereas doubling times could not be calculated for the strains that displayed extremely poor growth. These data strongly suggested that the absence of *cbiP* function correlated with the predicted phenotype of a strain lacking AdoCby synthase activity under conditions that demand de novo synthesis of cobamides. This finding led to the proposal that ORF Vng1576G was the archaeal ortholog of the CbiP.

***Halobacterium* can salvage Cbi.** Having a *Halobacterium* mutant blocked before the late steps of cobamide biosynthesis allowed us to test if this archaeon can salvage Cbi. In bacteria, AdoCbi is not an intermediate of the de novo pathway (8, 36, 39) (Fig. 1), and it is also not predicted to be an intermediate in archaea, based on the presence of CbiR. The salvaging of Cbi, therefore, would require additional enzymes or functions. The addition of Cbi to the medium allowed wild-type growth (i.e., 24-h doubling time) of strain JE6738 (Δ*cbiP*) (Fig. 3A) but did not significantly enhance the growth of the wild-type strain (data not shown). The ability of *Halobacterium* to salvage Cbi suggested the existence of an enzyme that can convert Cbi to a true intermediate of the de novo pathway. A mutation in the CbiB enzyme would block the pathway at a point that would allow us to ascertain whether the entry point for Cbi salvaging in archaea occurred via AdoCbi-P (as in bacteria) or via a new metabolic route.

***cbiB* (ORF Vng1578H) is a cobamide biosynthetic gene in *Halobacterium*.** Unlike strain MPK414, strain JE6791 (Δ*cbiB*) cannot grow in the defined medium lacking corrinoids (Fig. 3B). To test if the lack of growth was due to the inability to synthesize cobamides, Cbi-GDP (a pathway intermediate downstream of the CbiB-catalyzed reaction) (Fig. 1) was added to the medium. Cbi-GDP restored the growth of strain JE6791 (30-h doubling time) (Fig. 3B) but did not significantly enhance growth of the wild-type strain MPK414 (data not shown). The addition of Cby (a pathway intermediate prior to the CbiB catalyzed reaction), however, failed to restore growth of strain JE6791 (Fig. 3B). These results were consistent with a block in the synthesis of AdoCbi-P and led us to propose that ORF Vng1578H in *Halobacterium* encodes the archaeal ortholog of *S. enterica* CbiB enzyme.

CbiB activity is required for Cbi salvaging. As mentioned above, strain JE6738 (Δ*cbiP*) can salvage Cbi; however, the addition of Cbi to the medium did not restore the growth of strain JE6791 (Δ*cbiB*) (Fig. 3B). These results confirmed that in *Halobacterium* Cbi must enter the de novo pathway at an entry point prior to the CbiB-catalyzed step. This finding is also consistent with the observation that Cbl and AdoCbi are not intermediates of the archaeal de novo pathway. If they were, strain JE6791 would be predicted to be able to salvage Cbl.

Complementation of *cbiP* and *cbiB* mutants of *Halobacterium*. The observed AdoCby auxotrophy of JE6738 (Δ*cbiP*) and the AdoCbi-GDP auxotrophy of JE6791 (Δ*cbiB*) were corrected when the *cbiP*⁺ and *cbiB*⁺ alleles were reintroduced into the appropriate strains. Strain JE7001 (Δ*cbiP* ura3::*cbiP*⁺) and strain JE6930 (*cbiB*⁺ ura3::*cbiB*⁺) grew in the defined medium without any corrinoid supplementation (Fig. 3) with a doubling time of 26 and 34 h, respectively. The growth rate of these strains was similar to the rates of strains JE6738 (Δ*cbiP*) and JE6791 (Δ*cbiB*) growing on medium supplemented with the correct corrinoid supplements. These results showed that the *cbiP*⁺ or *cbiB*⁺ functions were necessary and sufficient to restore de novo cobamide synthesis in the mutant strains.

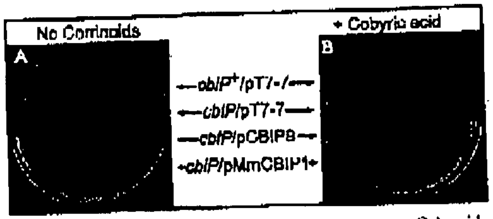


FIG. 4 Nutritional studies of *S. enterica cbiP* mutants. Cobamide-dependent growth of *S. enterica* strains grown anaerobically in defined solid medium at 37°C without a corrinoid supplement (A) or with 15 nM $(CN)_2$Cby (B). The strains are indicated by their genotypes. The strains used were TR6583 *metE cbiP*⁺ and JE588 *metE cbiP*. The plasmids used were pT7-7, vector-only control; pCBIP9, *S. enterica cbiP*⁺; and pMmCBIP1, *M. mazei cbiP*⁺.

The archaeal *cbiP* and *cbiB* genes complement *S. enterica cbiP* and *cbiB* mutants. To further support the conclusion that the archaeal orthologs of *cbiP* and *cbiB* do function as AdoCby and AdoCbi-P synthases in vivo, we tested the ability of archaeal *cbiP* and *cbiB* orthologs to complement *S. enterica cbiP* and *cbiB* mutants. To investigate this possibility, the *cbiP* and *cbiB* orthologs from the archaeal methanogen *M. mazei* strain Goe1 were cloned. Previous work in the laboratory has shown that *Halobacterium* genes do not express well in *S. enterica*, whereas genes from archaeal methanogens are well expressed (36). *M. mazei* ORF Mm0093 (gi number 21226195) showed 42% identity and 58% similarity to the *Halobacterium cbiP* gene, and ORF Mm2059 (gi number 21228161) showed 28% identity and 45% similarity to the *cbiB* gene of *Halobacterium*.

For this purpose, *S. enterica* strains carrying null alleles of *metE* and either *cbiP* or *cbiB* were used. The mutation in *metE* inactivates the cobamide independent methionine synthase (MetE) enzyme, thus demanding cobamide-dependent methylation of homocysteine to yield methionine by the action of the MetH enzyme (35). An insertion in either *cbiP* or *cbiB* eliminated de novo cobamide synthesis.

For *cbiP* complementation, the positive control plasmid pCBIP9 (containing a wild-type allele of *S. enterica cbiP*⁺) or plasmid pMmCBIP1 (*M. mazei cbiP*⁺) was introduced into the *S. enterica cbiP metE* mutant strain JE588.

For *cbiB* complementation, a plasmid containing a wild-type allele of either *S. enterica cbiB* (the positive control plasmid pSeCBIB4) or *M. mazei cbiB* (plasmid pMmCBIB1) was introduced into the *S. enterica cbiB metE* mutant strain JE6368. Residual expression of the *cbiP* or *cbiB* genes in the absence of the T7 RNA polymerase allowed us to assess complementation. In both cases, plasmid pT7-7 was used as a vector-only negative control.

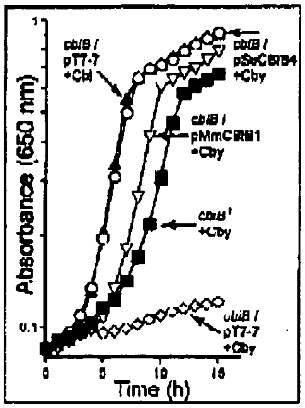


FIG. 5. Nutritional studies of *S. enterica cbiB* mutants. Cobamide-dependent growth of *S. enterica* strains grown aerobically in defined liquid medium at 37°C is reported as absorbance at 650 nm as a function of time. The strains are indicated by their genotypes. The corrinoids added to the medium are indicated next to the genotypes. The strains used were TR6583 *metE cbiP*+ and JE6368 *metE cbiB*. The plasmids used were pT7-7, vector control; pSeCBIB4, *S. enterica cbiB*+; and pMmCBIB1, *M. mazei cbiB*+. Abbreviations: Cby, cobyric acid dicyanide; Cbi, cobinamide dicyanide. In all cases, corrinoids were used at concentrations of 15 nM.

To test *cbiP* complementation, *S. enterica* was grown anaerobically, where the cells can synthesize cobamides de novo. Complementation of cobamide biosynthesis was observed when either *S. enterica* or *M. mazei cbiP* was provided in *trans* to JE588 but not with the control vector (Fig. 4A). Growth was similar for all strains when (CN)$_2$Cby was added (Fig. 4B). These results were consistent with the archaeal CbiP enzyme having AdoCby synthase activity in vivo.

cbiB complementation was tested under aerobic conditions, where *S. enterica* must salvage cobamide precursors. In this case Cby was added to the medium. Cby salvaging requires a functional CbiB synthase enzyme (Fig. 1); hence, growth on this intermediate would indicate restoration of the de novo pathway of *cbiB* mutant strain JE6368. Complementation of Cby salvaging was observed when either *S. enterica cbiB* (pSeCBIB4) or *M. mazei cbiB* (pMmCBIB1) was provided in *trans* but not when the control vector was provided (Fig. 5). These data support the conclusion that the archaeal CbiB enzyme has AdoCbi-P synthase activity in vivo.

DISCUSSION

The contributions of this work are twofold. First, the functions encoded by two putative ORFs in two archaea are supported by in vivo evidence. Second, evidence for the existence of the pathway for salvaging the cobamide precursor Cbi in archaea has been obtained. The latter pathway is distinct from the one used by bacteria.

Biochemical roles of two archaeal genes in cobamide biosynthesis. The results of the nutritional analysis of mutants of the extremely halophilic archaeon *Halobacterium* sp. strain NRC-1 showed that ORFs Vng1576G and Vng1578H were necessary for de novo cobamide biosynthesis and that ORF Vng1578H was necessary for salvaging cobyric acid from the environment. The conclusions drawn from these analyses were fully supported by complementation analyses of bona fide *S. enterica* mutants lacking either CbiP or CbiB activities by *M. mazei* strain Goe1 genes. On the basis of this work, we propose that *Halobacterium* ORF Vng1578H be annotated as encoding the AdoCbi-P synthase enzyme and that the putative annotation of Vng1576G as encoding the AdoCby synthase enzyme is correct. ORF Vng1578H should be renamed as *cbiB* to reflect its involvement in cobamide biosynthesis in archaea. This nomenclature should be extended to the ORFs Mm0093 (*cbiP*) and Mm2059 (*cbiB*) of *M. mazei* strain Goe1.

In this study, corrinoid intermediates have been assumed to be adenosylated in vivo. Although this fact has been established in bacteria (12), it is unknown if the corrinoids are adenosylated in archaea. Because archaea possess a putative

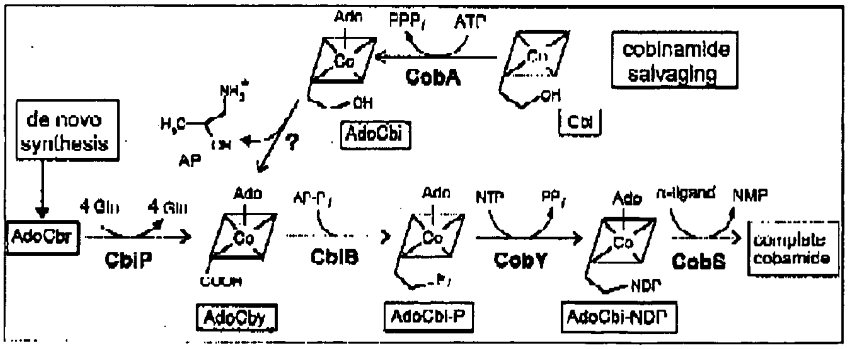


FIG. 6. A new model for the late steps of cobamide biosynthesis in archaea. Intermediates are boxed and indicated below structures. The adenosylation of archaeal intermediates is putative. The putative archaeal orthologs of the bacterial CobA and CobS (16) proteins are indicated. Abbreviations: AP-P$_i$, aminopropanol phosphate; AP, aminopropanol; AdoCbr, adenosylcobyrinic acid a,c-diamide; AdoCby, adenosylcobyric acid; AdoCbi, adenosylcobinamide; CobS, cobalamin (5′-P) synthase; CobY, NTP:AdoCbi-P nucleotidyltransferase.

ortholog of CobA and archaeal genes can complement *S. enterica cob* mutants, it is assumed that the corrinoid substrates for the archaeal enzymes are adenosylated.

The archaeal pathway for salvaging Cbi is different from the bacterial pathway. The requirement for CbiB enzyme activity for the salvaging of Cbi by *Halobacterium* is key to the proposal that the archaeal pathway for salvaging this precursor is different from the one that operates in bacteria (Fig. 1 and 6). In bacteria, CbiB is not required for Cbi salvaging because the NTP:AdoCbi kinase activity of CobU directly converts AdoCbi to AdoCbi-P, the product of the CbiB enzyme (Fig. 1). The kinase activity of CobU effectively bypasses the need for CbiB. The tight block in Cbi salvaging observed in *Halobacterium cbiB* mutants strongly suggests that the point of entry of Cbi salvaging in this archaeon is AdoCby, which can then be converted by the action of CbiB to AdoCbi-P, the substrate for the next enzyme of the archaeal pathway, i.e., CobY (Fig. 6). It is unlikely that the point of entry is prior to AdoCby, because *Halobacterium cbiP* mutants can readily salvage Cbi. We propose that, in archaea, AdoCbi is the substrate for an unidentified amidohydrolase enzyme that cleaves off the (*R*)-1-amino-2-propanol moiety of AdoCbi to yield AdoCby, the substrate of CbiB (Fig. 6). We favor this hypothesis on the basis of preliminary data obtained in our laboratory, which show that this AdoCbi amidohydrolase activity is present in cell extracts of *E. coli* overexpressing a single gene of *M. mazei* (J. D. Woodson and J. C. Escalante-Semerena, unpublished results). The requirement of an adenosylated substrate is speculative, and it is possible that the corrin ring is adenosylated after entering the de novo pathway. The identification of the gene encoding the amidohydrolase activity and the isolation and characterization of this new cobamide biosynthetic enzyme will be reported elsewhere.

ACKNOWLEDGMENTS

This work was supported by a grant GM40313 from the National Institutes of Health (NIH) to J.C.E.-S.; J.D.W. was supported in part by the Ira L. Baldwin Predoctoral Fellowship; C.L.Z. was supported in part by NIH Minority Access to Research Careers Predoctoral Fellowship F31-GM64009.

We thank P. Renz for his gift of $(CN)_2Cby$ and G. Gottschalk for his gift of *M. mazei* chromosomal DNA.

REFERENCES

1. Balch, W. E., and R. S. Wolfe. 1976. New approach to the cultivation of methanogenic bacteria: 2-mercaptoethanesulfonic acid (HS-CoM)-dependent growth of *Methanobacterium ruminantium* in a pressurized atmosphere. Appl. Environ. Microbiol. 32:781–791.
2. Battersby, A. R. 2000. Tetrapyrroles: the pigments of life. Nat. Prod. Rep. 17:507–526.
3. Berkowitz, D., J. M. Hushon, H. J. Whitfield, J. Roth, and B. N. Ames. 1968. Procedure for identifying nonsense mutations. J. Bacteriol. 96:215–220.
4. Blanche, F., B. Cameron, J. Crouzet, L. Debussche, D. Thibaut, M. Vuilhorgne, F. J. Leeper, and A. R. Battersby. 1995. Vitamin B_{12}: how the problem of its biosynthesis was solved. Angew. Chem. Int. Ed. Engl. 34:383–411.
5. Blanche, F., L. Debussche, A. Famechon, D. Thibaut, B. Cameron, and J. Crouzet. 1991. A bifunctional protein from *Pseudomonas denitrificans* carries cobinamide kinase and cobinamide phosphate guanylyltransferase activities. J. Bacteriol. 173:6052–6057.
6. Brey, R. N., C. D. B. Banner, and J. B. Wolf. 1986. Cloning of multiple genes involved with cobalamin (vitamin B_{12}) biosynthesis in *Bacillus megaterium*. J. Bacteriol. 167:623–630.
7. Brindley, A. A., E. Raux, H. K. Leech, H. L. Schubert, and M. J. Warren. 2003. A story of chelatase evolution: identification and characterization of a small 13–15-kDa "ancestral" cobaltochelatase (CbiXS) in the archaea. J. Biol. Chem. 278:22388–22395.
8. Brushaber, K. R., G. A. O'Toole, and J. C. Escalante-Semerena. 1998. CobD, a novel enzyme with L-threonine-O-3-phosphate decarboxylase activity, is responsible for the synthesis of (*R*)-1-amino-2-propanol O-2-phosphate, a proposed new intermediate in cobalamin biosynthesis in *Salmonella typhimurium* LT2. J. Biol. Chem. 273:2684–2691.
9. DasSarma, S., U. L. RajBhandary, and H. G. Khorana. 1984. Bacterio-opsin mRNA in wild-type and bacterio-opsin-deficient *Halobacterium halobium* strains. Proc. Natl. Acad. Sci. USA 81:125–129.
10. DiMarco, A. A., T. A. Bobik, and R. S. Wolfe. 1990. Unusual coenzymes of methanogenesis. Annu. Rev. Biochem. 59:355–394.
11. Escalante-Semerena, J. C. 1999. Regulation of adenosylcobalamin biosynthesis in *Salmonella typhimurium*, p. 577–594. *In* R. Banerjee (ed.), Chemistry and biochemistry of B_{12}. Academic Press, Inc., New York, N.Y.
12. Escalante-Semerena, J. C., S.-J. Suh, and J. R. Roth. 1990. *cobA* function is required for both de novo cobalamin biosynthesis and assimilation of exogenous corrinoids in *Salmonella typhimurium*. J. Bacteriol. 172:273–280.
13. Friedmann, H. C., and R. K. Thauer (ed.). 1992. Macrocyclic tetrapyrrole biosynthesis in bacteria, vol. 3. Academic Press, Inc., New York, N.Y.
14. Grey, V. L., and P. S. Fitt. 1976. An improved synthetic growth medium for *Halobacterium cutirubrum*. Can. J. Microbiol. 22:440–442.
15. Krebs, M. P., R. Mollaaghababa, and H. G. Khorana. 1993. Gene replacement in *Halobacterium halobium* and expression of bacteriorhodopsin mutants. Proc. Natl. Acad. Sci. USA 90:1987–1991.
16. Maggio-Hall, L. A. 2001. Synthesis and incorporation of the lower ligand base of cobalamin. Ph.D. dissertation. University of Wisconsin, Madison.
17. Ng, W. V., S. P. Kennedy, G. G. Mahairas, B. Berquist, M. Pan, H. D. Shukla, S. R. Lasky, N. S. Baliga, V. Thorsson, J. Sbrogna, S. Swartzell, D. Weir, J. Hall, T. A. Dahl, R. Welti, Y. A. Goo, B. Leithauser, K. Keller, R. Cruz, M. J. Danson, D. W. Hough, D. G. Maddocks, P. E. Jablonski, M. P. Krebs, C. M. Angevine, H. Dale, T. A. Isenbarger, R. F. Peck, M. Pohlschroder, J. L. Spudich, K. W. Jung, M. Alam, T. Freitas, S. Hou, C. J. Daniels, P. P. Dennis, A. D. Omer, H. Ebhardt, T. M. Lowe, P. Liang, M. Riley, L. Hood, and S. DasSarma. 2000. Genome sequence of *Halobacterium* species NRC-1. Proc. Natl. Acad. Sci. USA 97:12176–12181.
18. Oesterhelt, D., and W. Stoeckenius. 1974. Isolation of the cell membrane of *Halobacterium halobium* and its fractionation into red and purple membrane. Methods Enzymol. 31:667–678.
19. O'Toole, G. A., and J. C. Escalante-Semerena. 1995. Purification and characterization of the bifunctional CobU enzyme of *Salmonella typhimurium* LT2: evidence for a CobU-GMP intermediate. J. Biol. Chem. 270:23560–23569.
20. Peck, R. F., S. DasSarma, and M. P. Krebs. 2000. Homologous gene knockout in the archaeon *Halobacterium salinarum* with ura3 as a counterselectable marker. Mol. Microbiol. 35:667–676.
21. Peck, R. F., C. Echavarri-Erasun, E. A. Johnson, W. V. Ng, S. P. Kennedy, L. Hood, S. DasSarma, and M. P. Krebs. 2001. *brp* and *blh* are required for synthesis of the retinal cofactor of bacteriorhodopsin in *Halobacterium salinarum*. J. Biol. Chem. 276:5739–5744.
22. Peck, R. F., E. A. Johnson, and M. P. Krebs. 2002. Identification of a lycopene beta-cyclase required for bacteriorhodopsin biogenesis in the archaeon *Halobacterium salinarum*. J. Bacteriol. 184:2889–2897.
23. Raux, E., A. Lanois, A. Rambach, M. J. Warren, and C. Thermes. 1998. Cobalamin (vitamin B_{12}) biosynthesis: functional characterization of the *Bacillus megaterium cbi* genes required to convert uroporphyrinogen III into cobyrinic acid *a*, *c* diamide. Biochem. J. 335:167–173.
24. Raux, E., A. Lanois, M. J. Warren, A. Rambach, and C. Thermes. 1998. Cobalamin (vitamin B_{12}) biosynthesis: identification and characterization of a *Bacillus megaterium cobI* operon. Biochem. J. 335:159–166.
25. Riera, J., F. T. Robb, R. Weiss, and M. Fontecave. 1997. Ribonucleotide reductase in the archaeon *Pyrococcus furiosus*: a critical enzyme in the evolution of DNA genomes? Proc. Natl. Acad. Sci. USA 94:475–478.
26. Roessner, C. A., P. J. Santander, and A. I. Scott. 2001. Multiple biosynthetic pathways for vitamin B12: variations on a central theme. Vitam. Horm. 61:267–297.
27. Rondon, M. R., J. R. Trzebiatowski, and J. C. Escalante-Semerena. 1997. Biochemistry and molecular genetics of cobalamin biosynthesis. Prog. Nucleic Acid Res. Mol. Biol. 56:347–384.
28. Roth, J. R., J. G. Lawrence, and T. A. Bobik. 1996. Cobalamin (coenzyme B_{12}): synthesis and biological significance. Annu. Rev. Microbiol. 50:137–181.
29. Santander, P. J., C. A. Roessner, N. Stolowich, M. T. Holderman, and A. I. Scott. 1997. How corrinoids are synthesized without oxygen: nature's first pathway to vitamin B_{12}. Chem. Biol. 4:659–666.
30. Scott, A. I. 2003. Discovering nature's diverse pathways to vitamin B_{12}: a 35-year odyssey. J. Org. Chem. 68:2529–2539.
31. Scott, A. I., and C. A. Roessner. 2002. Biosynthesis of cobalamin (vitamin B_{12}). Biochem. Soc. Trans. 30:613–620.
32. Stupperich, E., I. Steiner, and H. J. Eisinger. 1987. Substitution of Coα-(5-hydroxybenzimidazolyl)cobamide (factor III) by vitamin B_{12} in *Methanobacterium thermoautotrophicum*. J. Bacteriol. 169:3076–3081.
33. Tabor, S. 1990. Expression using the T7 RNA polymerase/promoter system.

p. 16.2.1–16.2.11. *In* F. M. Ausubel, R. Brent, R. E. Kingston, D. D. Moore, J. G. Seidman, J. A. Smith, and K. Struhl (ed.), Current protocols in molecular biology, vol. 2. Wiley Interscience, New York, N.Y.

34. Tauer, A., and S. A. Benner. 1997. The B_{12}-dependent ribonucleotide reductase from the archaebacterium *Thermoplasma acidophila*: an evolutionary solution to the ribonucleotide reductase conundrum. Proc. Natl. Acad. Sci. USA 94:53–58.

35. Taylor, R. T., and H. Weissbach. 1973. N^5-methyltetrahydrofolate-homocysteine methyltransferases, p. 121–165. *In* P. D. Boyer (ed.), The enzymes, vol. 9. Academic Press, Inc., New York, N.Y.

36. Thomas, M. G., and J. C. Escalante-Semerena. 2000. Identification of an alternative nucleoside triphosphate: 5′-deoxyadenosylcobinamide phosphate nucleotidyltransferase in *Methanobacterium thermoautotrophicum* ΔH. J. Bacteriol. 182:4227–4233.

37. Tsai, S. P., R. J. Hartin, and J. Ryu. 1989. Transformation in restriction-deficient *Salmonella typhimurium* LT2. J. Gen. Microbiol. 135:2561–2567.

38. Warren, M. J., E. Raux, H. L. Schubert, and J. C. Escalante-Semerena. 2002. The biosynthesis of adenosylcobalamin (vitamin B_{12}). Nat. Prod. Rep. 19:390–412.

39. Woodson, J. D., R. F. Peck, M. P. Krebs, and J. C. Escalante-Semerena. 2003. The *cobY* gene of the archaeon *Halobacterium* sp. strain NRC-1 is required for de novo cobamide synthesis. J. Bacteriol. 185:311–316.

This Page is Inserted by IFW Indexing and Scanning Operations and is not part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

☐ BLACK BORDERS

☐ IMAGE CUT OFF AT TOP, BOTTOM OR SIDES

☐ FADED TEXT OR DRAWING

☐ BLURRED OR ILLEGIBLE TEXT OR DRAWING

☐ SKEWED/SLANTED IMAGES

☐ COLOR OR BLACK AND WHITE PHOTOGRAPHS

☐ GRAY SCALE DOCUMENTS

☐ LINES OR MARKS ON ORIGINAL DOCUMENT

☑ REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY

☐ OTHER: ______________________________

IMAGES ARE BEST AVAILABLE COPY.
As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.